AUG 25

VOL. XXIX MAY, 1938 NO. 5

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

Vocational Attitude Patterns in High-school Students 321
HAROLD D. CARTER AND MARY COVER JONES

A Study of Clinical Predictions of Student Success or Failure in Professional Training 335
JOHN G. DARLEY

The Estimation of Specific and Bi-factors 355
GODFREY H. THOMSON

The Relationship between Scores on the Scholastic Aptitude Test and College Grades in Mathematics 363
M. RICHARD DICKTER

Rosanoff, Handy, and Plesset on the Etiology of Mental Deficiency: A Critical Appraisal 374
FRANK K. SHUTTLEWORTH

A Statistic for Circular Scales 384
ROBERT T. ROSS

Intelligence Testing by Drawings 390
JOSEPH MILLER

Book Reviews 395

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
University of Rochester

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIX May, 1938 Number 5

VOCATIONAL ATTITUDE PATTERNS IN HIGH-SCHOOL STUDENTS

HAROLD D. CARTER AND MARY COVER JONES*

Institute of Child Welfare, University of California

I. INTRODUCTION

This paper reports results from application of the Strong Vocational Interest Blank to a group of high-school students. This test was developed for use with older persons, and it has been normed through study of men selected from a number of occupational groups. The present extension of the use of the inventory is justified by two considerations: First, recent evidence[2,3,4,6] indicates that vocational attitudes as measured by the test are present at an earlier age than was formerly supposed; second, the need of objective techniques for use by high-school counselors makes it desirable to continue research upon vocational attitudes of high-school boys and girls.

In studying the developing vocational attitudes of high-school students as measured by the inventory technique, it is necessary at the present time to deal largely with basic problems and to correlate the results of measurement with other facts about the children. Hence a number of advantages in the present study result from its incorporation in a larger investigation. The boys and girls examined here are those included in the University of California Adolescent Study. The inventory results are, therefore, supplemented by much additional data, including subjective insights gained from long experience with the individual children. The present report is a preliminary explora-

* The collection of the data was made possible through the coöperation of the Oakland Public Schools, and the tabulations and computations were facilitated by W.P.A. Project Number 4428. The writers are indebted to Herbert S. Conrad and Daniel A. Prescott, who read and criticised the manuscript, and to Harold E. Jones for administrative assistance and valuable criticism.

tion of data from the first series of tests in a projected seriatim study of the sampling.

II. THE LITERATURE

A great many studies have been made of the occupational choices of high-school students; except for recent contributions these have been summarized in Fryer's 1931 book.[7] The published accounts suggest that the vocational choices of high-school students are unstable and unreliable, that they often lack practical significance because they are autistic, and that in only a small percentage of cases are the choices valid indications of occupational attitude. However, these views are based largely upon studies conducted by the questionnaire method, unsupported by objective test data, and suffering from the faults usually associated with the direct questionnaire approach.

Studies using objective tests have been somewhat more promising, although the long history of interest measurement has been replete with only partially successful experiments in test construction. The most successful test technique has been that of the interest inventory. Of the available inventories, the Strong Vocational Interest Blank has been selected for use in this study, because it has been most widely standardized, and because it has the unique and practical advantage of occupational norms. Studies with this test have indicated that the traits measured show but slight changes with age.[2,4,15,20] Allowance for these changes can be made, because they involve more or less predictable trends. Using the actual vocational choices and shifts of occupation of adults as a criterion, recent studies[12,18,19] have shown that the test scores have considerable validity, although the evidence is attenuated by unreliability of the criteria. Evidence is available to show that the test scores are highly reliable.[18,19] Studies[4,9,12,14,20] by various techniques, including that of factor analysis, have indicated that the test measures four or more independent constellations of attitude, and that occupations can be classified into a small number of family groups on the basis of these interests or attitudes. The data concerning the origin and development of these attitudes are limited, but the evidence[2,3,4,6,7,8,9] indicates that the attitudes are present before entrance into an occupation, and before entering upon a course of formal training for an occupation.

It is recognized that the test has limited usefulness for lower occupational classifications, but this defect may be corrected in the light of present research. Moreover, the defect is not so serious as

one might at first think, because the main need is for interest-testing devices in connection with professional and skilled groups. It is in these latter cases that the counselor most needs objective aids, since choices here involve long courses of training, necessitating extensive investment of energy, money, and time.

III. THE DATA

The Strong Vocational Interest Blank for men* was administered to a group of two hundred eight high-school students, including seventy-eight boys and one hundred thirty girls. This sampling included an intensively-studied group of eighty-three individuals, referred to in this paper as Group *A*. Blanks filled out by members of this group were scored by means of at least ten of the Strong Interest Scales, and the scores have been studied in relation to several other variables. The blanks filled out by an additional one hundred twenty-five children, herein referred to as Group *B*, were scored for the single occupations chosen by them. The data presented below indicate the age and sex composition of each group, and certain relevant facts concerning their scores on the inventory.

The sampling is relatively homogeneous in age and educational attainment. In Group *A*, the forty-four boys have a mean age of 15.7 years, with a standard deviation of .50 years; the forty girls have a mean age of 15.6 years, with a standard deviation of .58 years. The range of ages is 14.6 to 17.2 for the boys, and 14.8 to 16.8 for the girls. All these children were in the tenth grade of a public school in Oakland, California, at the time of testing. The additional one hundred twenty-five students in Group *B* are classmates of the students in Group *A*.

In the study of each inventory filled out by a member of Group *A*, the following scales for the Strong Vocational Interest Blank were used: Physicist (and Chemist), Lawyer, Salesman (Real Estate and Life Insurance), Teacher, Secretary, Office Worker, and Certified Public Accountant. These scales are considered to be diagnostic of seven different types of interest, as pointed out by Strong in the *Manual*.[12] In addition to this, each blank was scored for the indi-

* In administering the men's blank to the girls, we have several purposes in mind. The men's blank is better standardized in some respects, and contains many of the items also included in the women's blank. Although it is probably not advisable to use the men's blank routinely in work with girls, it has, nevertheless, seemed to us a desirable procedure in a research in which sex differences are being studied.

vidual's chosen occupation wherever possible; this additional scoring required the following scales: Advertiser, Architect, Artist, Doctor, Engineer, Journalist, Minister, Musician, Personnel Manager, and Purchasing Agent. Finally, the blanks were routinely scored, using Strong's Interest Maturity Scale, Carter's Masculinity Scale,[5] and the Young-Estabrooks Studiousness Scale. In all, this includes twenty-two scales of the interest inventory which were employed in the study of the primary group.* Each blank filled out by a member of Group *B* was scored by means of the most appropriate single scale from the above list.

IV. PROCEDURE

Since the Strong Vocational Interest Blank was originally intended for adult males, its application to this group of high-school students is not quite a routine matter. It is, rather, a research problem, requiring new procedures to permit evaluation and interpretation of the test results. Changes in scores with age, and sex differences in scores, are recognized as factors requiring special treatment.

The inventory employed here had never been intended for use with women and girls, hence adequate explanation of the girls' scores involves research and analysis beyond the scope of the present paper. We have administered the test to the girls in our sample in order to get data for direct comparison of girls and boys. It also seems likely that the results will supplement those obtained later by means of the Women's Blank, especially since earlier research[3] has indicated that the interests of girls are not confined to those occupations traditionally open to women. The women's blank has many items in common with those in the men's blank, but the occupational norms probably have different significance for the two sexes. We have tabulated the stated occupational choices of the girls and have scored their blanks, using the appropriate scales where possible, and present the results along with similar results for boys. This procedure has furnished interesting comparisons. But we are not here attempting to interpret and evaluate the results for the girls.

In regard to age differences, the fundamental need felt here is for high-school norms, and for seriatim study to show the developmental significance of the scores. Both these problems are being attacked. Research has shown[2,4,15] that the changes in scores with age are gradual

* Persons using the Strong Vocational Interest Blank may purchase the scoring scales from the Stanford University Press.

and more or less predictable, and there is every reason to suppose that the scores of high-school students represent an orderly arrangement of attitudes, at a slightly less well-developed stage than that found among college students.

In the present preliminary study, the scores are treated in two ways, in the attempt to interpret the results. First, the ratings based on Strong's adult norms[12] are used as an indication of degree of development toward adult standards. This is a severe criterion, but it has practical significance: The *A* rating means "Yes, you do have the pattern of interests characteristic of persons successfully engaged in the given occupation." The *C* rating means "No, you do not have the pattern of interests characteristic of persons in that occupation." The *B* ratings are intermediate. It is probable that in some instances high-school students tend to receive *B* ratings instead of *A* ratings because of their relative immaturity.

As a second technique, standard scores have been employed in order to examine the individuals with reference to the distribution of scores for a high-school population. These standard scores indicate the position of the individual among others of his own age and grade; hence they furnish a significant indication of the meaning of his scores. In the present study, standard scores have been used extensively; for convenience in statistical work they have been so arranged that the average score is 50, the standard deviation is ten points, and the range of scores is from 20 to 80. This arrangement enables us to place an individual in his position in the group at a glance, irrespective of the scale employed.

These data have been used in order to answer a number of questions. First, do the scores on the test bear some relationship to the stated vocational choices of the students, and, if so, what? Second, what are some of the most significant facts about age, intelligence, and other variables in relation to vocational choices? Third, does the study of profiles of scores of individuals indicate the presence of an organized pattern of interests? Finally, does a preliminary study of this sort make possible recommendations concerning the use of the test at the high-school level?

V. RESULTS

The first question to be approached here concerns the relationship between test score and the stated occupational choices of the individuals. In Table I is given a summary of the results based upon the

total group of one hundred seventy-one students who made a choice. The results of this summary show conclusively that when scored on the occupational scales most appropriate to their choices the high-

TABLE I.—RATINGS OF TENTH-GRADE STUDENTS ON INTEREST INVENTORY SCALES APPROPRIATE TO THEIR OCCUPATIONAL CHOICES

Ratings	67 boys		104 girls		Total 171	
	n	Per cent	*n*	Per cent	*n*	Per cent
A	14	20.9	15	14.4	29	17.0
B+	13	19.4	15	14.4	28	16.4
B	16	23.9	20	19.2	36	21.0
B−	6	9.0	12	11.5	18	10.5
C	18	26.9	42	40.4	60	35.1
Total	67	100.0	104	99.9	171	100.0
Above *C*	49	73.2	62	59.6	111	64.9

school boys and girls secure a high proportion of ratings above *C*. There is a goodly proportion of *A* ratings. In only a minority are the scores sufficiently low to indicate definitely an absence of the pattern of interests appropriate to the chosen occupational groups.

In the intensively-studied group, forty-six students indicated a definite choice of occupation. Table II shows the summary of ratings received by members of this subsample. The results are essentially similar to those in Table I. This second table is necessary for comple-

TABLE II.—RATINGS OF TENTH-GRADE STUDENTS ON INTEREST INVENTORY SCALES APPROPRIATE TO THEIR OCCUPATIONAL CHOICES. DATA FROM A SUBSAMPLE OF FORTY-SIX STUDENTS INCLUDED IN AN INTENSIVE STUDY PROGRAM

Ratings	17 boys		29 girls		Total 46	
	n	Per cent	*n*	Per cent	*n*	Per cent
A	3	17.6	5	17.2	8	17.4
B+	4	23.5	2	6.9	6	13.0
B	3	17.6	4	13.8	7	15.2
B−	3	17.6	6	20.7	9	19.6
C	4	23.5	12	41.4	16	34.8
Total	17	100.0	29	100.0	46	100.0
Above *C*	13	76.3	17	58.6	30	65.2

tion of the argument, because in further comparisons only this particular subsample is represented.

Since each person's blank was scored on each of seven scales, in the routine treatment each person has received seven ratings, no matter what his choice of occupation may be. We have here, then, the distribution of ratings on a representative sampling of the interest scales. These data for the same forty-six individuals represented in Table

TABLE III.—RATINGS OF TENTH-GRADE STUDENTS ON ARBITRARILY-SELECTED OCCUPATIONAL SCALES. DATA FROM FORTY-SIX STUDENTS IN THE INTENSIVE STUDY SAMPLE, WHO STATED AN OCCUPATIONAL CHOICE

Ratings	17 boys		29 girls		Total 46	
	n	Per cent	*n*	Per cent	*n*	Per cent
A	5	4.2	5	2.5	10	3.1
B+	16	13.4	27	13.3	43	13.4
B	9	7.6	31	15.3	40	12.4
B−	10	8.4	9	4.4	19	5.9
C	79	66.4	131	64.5	210	65.2
Total	119	100.0	203	100.0	322	100.0
Above *C*	40	33.6	72	35.5	112	34.8

II, are assembled in Table III. Here we see a striking reversal of the trends. When scored on arbitrarily-selected scales, without regard for their choices, these high-school boys and girls tend very definitely to secure low ratings. There are many *C* ratings, indicating a lack of the interests characterizing the occupational groups in question. *A* ratings are very rare. From comparison of Tables I, II, and III it is apparent that the scores on the Strong Blank tend to be much higher when the scales used are appropriate to the students' choices, and much lower when the scales are arbitrarily selected.

This comparison depends not on *one* arbitrarily-chosen scale, but upon seven. The results differ for different scales, but in this preliminary analysis it is convenient to limit the presentation to broad trends in the data.

Another comparison in this series will be instructive. In the intensively-studied group, thirty-seven students did not state a choice of occupation. Table IV shows the summary of ratings received by members of this subsample when their blanks were scored on the seven arbitrarily-chosen scales. The results of Table IV are similar to those

of Table III, but they indicate a still stronger trend toward low scores. The results harmonize with the theory that students who have no choice of occupation tend to have more poorly-developed patterns of interest.

TABLE IV.—RATINGS OF TENTH-GRADE STUDENTS ON ARBITRARILY-SELECTED OCCUPATIONAL SCALES. DATA FROM THIRTY-SEVEN STUDENTS IN THE INTENSIVE STUDY SAMPLE, WHO MADE NO CHOICE OF OCCUPATION

Ratings	26 boys		11 girls		Total 37	
	n	Per cent	*n*	Per cent	*n*	Per cent
A	4	2.2	0	0.0	4	1.5
B+	8	4.4	2	2.6	10	3.9
B	24	13.2	9	11.7	33	12.7
B−	11	6.0	8	10.4	19	7.3
C	135	74.2	58	75.3	193	74.5
Total	182	100.0	77	100.0	259	99.9
Above *C*	47	25.8	19	24.7	66	25.5

Detailed statistical comparisons between the groups represented in Tables III and IV were made; some of the findings are interesting in this connection. The forty-six students who stated an occupational choice were reliably brighter,* more mature, more studious, and less "masculine" than the thirty-seven who did not state a choice. The students who made a choice were also reliably younger, although the difference in chronological age seems objectively small and unimportant from a practical standpoint; this finding probably reflects the fact that in any school grade the superior students tend to be the younger ones.

The group who indicated a choice of occupation had higher scores on the Teacher, Salesman, Secretary, Office Worker, and C.P.A. scales. The no-choice group had higher scores on the Physicist and Lawyer scales. The only statistically reliable difference was in scores

* The intelligence of the group was measured by means of four tests, including the Stanford-Binet, the Kuhlmann-Anderson, the Terman Group Test, and the Thorndike CAVD. For the total group, the Stanford-Binet IQ's ranged from seventy-three to one hundred thirty-six, with a mean of one hundred four and a standard deviation of thirteen points. Each test shows the "Choice" group to be superior, but there is much overlapping. The composite of the four tests shows a statistically reliable difference.

on the C.P.A. scale; here the group who had indicated a choice of occupation made significantly higher scores. This finding fits in with the conclusions in an earlier paper[3] to the effect that maturity is associated with high scores on this scale.

No exhaustive study of sex differences is attempted in this paper, but the data in the foregoing tables reveal two tendencies. First, with respect to arbitrarily-selected scales, there is no large general difference between the sexes in the tendency to receive high ratings (See Tables III and IV). We know, however, that more detailed study will reveal sex differences in scores on particular scales. Second, boys tend more often than girls to receive high ratings on the scales appropriate to their occupational choices (See Tables I and II); this is partly due to the fact that the scales employed were somewhat *less* appropriate to certain of the girls' choices of occupation.

It seems apparent on the basis of experience that both the men's blank and the women's blank can profitably be used in study of the interests of girls. It is a matter of individual differences. Some girls' choices seem to require use of the scales for the men's blank; in other instances the men's blank seems quite inadequate.

VI. TREATMENT OF INDIVIDUALS

The second broad question approached in this preliminary study concerns the tendency of individuals to have well-organized patterns of interest at this age. This is a fundamental question about the development of occupational attitudes. The tendency to receive high ratings, as shown in the preceding tables, gives a partial answer. Further basic facts are revealed by study of profiles of standard scores.

Fig. 1 shows the scores of an individual boy, in profile form. This boy stated that engineering was his chosen occupation, and he received an A rating on the Engineer scale. This high-school sophomore has well-developed interests of the type characteristic of successful men in that occupation.

The profile presents at a glance the relevant data concerning this boy's vocational interest scores. His high score on the Physicist scale, his high IQ, his high score on the Masculinity scale, and his superior home environment all suggest that his stated vocational choice is suitable. The tendency for high scores on the Engineer scale to be associated with high scores on the Physicist scale and low scores on the Lawyer, Salesman, Teacher, Secretary, and Office Worker scales has been frequently noted in studying the data. The

tendency of engineering interests to be accompanied by low scores on the Interest Maturity and Studiousness scales has also been frequently

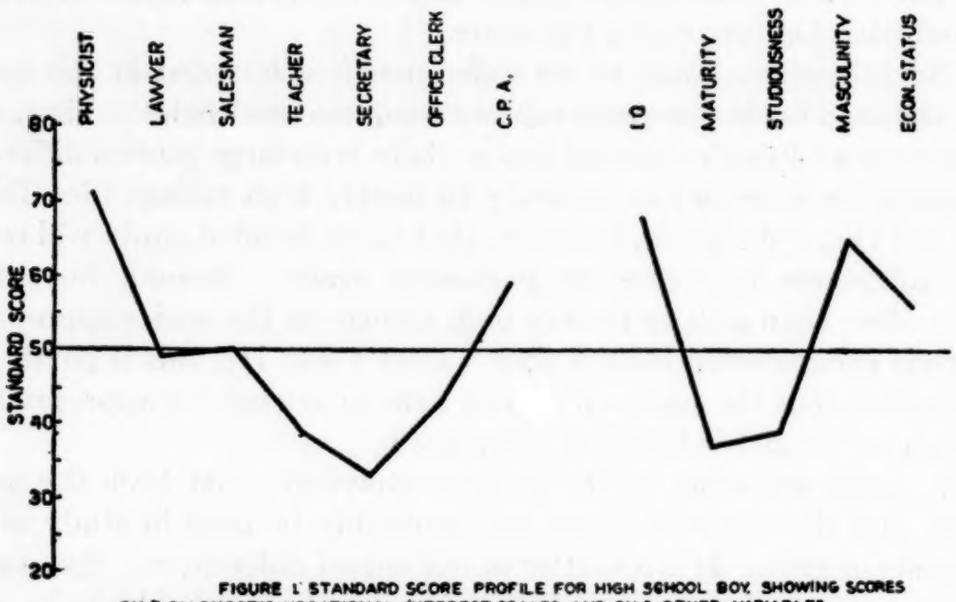

FIGURE 1. STANDARD SCORE PROFILE FOR HIGH SCHOOL BOY, SHOWING SCORES ON 7 DIAGNOSTIC VOCATIONAL INTEREST SCALES AND ON 5 OTHER VARIABLES.

noted. This boy has a definite *pattern* of interest scores, of a kind frequently found, and it is a pattern consonant with his choice of occupation.

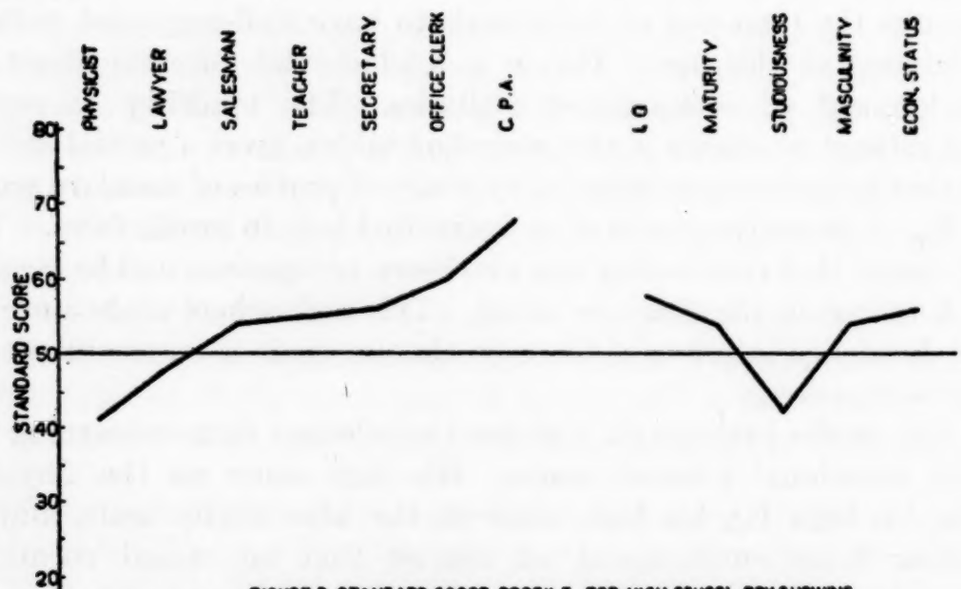

FIGURE 2. STANDARD SCORE PROFILE FOR HIGH SCHOOL BOY, SHOWING SCORES ON 7 DIAGNOSTIC VOCATIONAL INTEREST SCALES AND ON 5 OTHER VARIABLES.

Another pattern of interests is revealed in the profile for another boy, in Fig. 2. This boy stated that he wanted to "go into business";

he gave no more specific indication. His score on the Office Worker scale is $B+$, which is probably a favorable indication. The boy's scores are low on the Physicist scale and high on the Office Worker and C.P.A. scales. This latter finding is rare for high-school students. Since this boy has superior intelligence and is more mature than the average boy in his class, and since he comes from a home with superior financial advantages, it seems likely that he will find satisfaction in his chosen field of work. Although his choice is less definite or specific

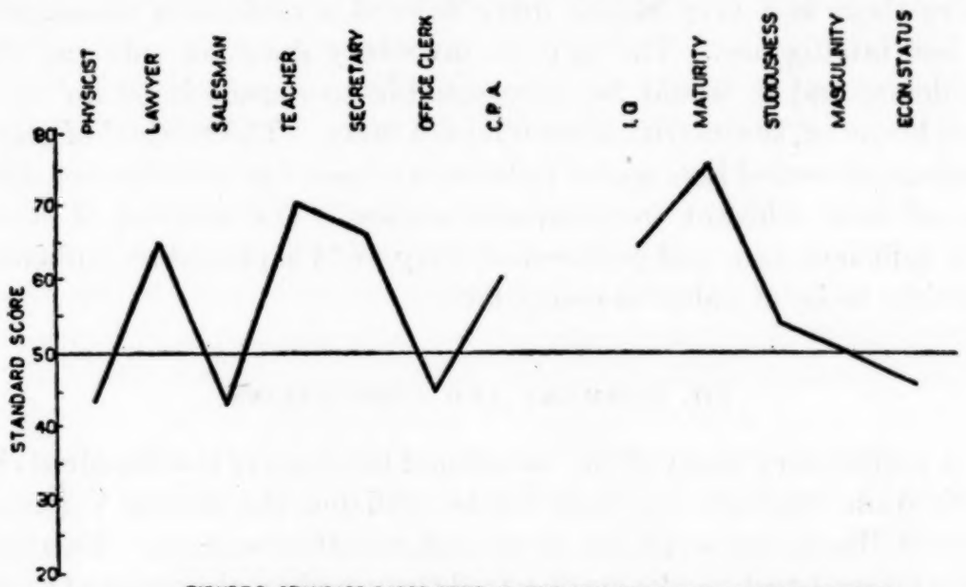

FIGURE 3. STANDARD SCORE PROFILE FOR HIGH SCHOOL BOY, SHOWING SCORES ON 7 DIAGNOSTIC VOCATIONAL INTEREST SCALES AND ON 5 OTHER VARIABLES.

than that of many, we note that he does have a pattern of interests in harmony with the general field of work indicated.

A third case is presented in Fig. 3. This is a boy who stated no choice of occupation. The test results help us to understand the difficulties which may be present, because his profile of interests suggests the possibility of conflict. There are several strongly-developed sets of attitudes, and there are other interests which are definitely lacking. This boy's intelligence score seems sufficiently high to satisfy the intelligence requirements of most occupations. He is very mature in relation to his classmates, and the moderate scores on the Studiousness and Masculinity scales are favorable indications. Since he comes from a home deprived of superior financial advantages, it may be wise to steer him away from Law, unless his drive is very strong. His highest scores are on the Teacher and Secretary scales. His problem is to make a choice among his interests; the counselor should probably furnish him with the evidence available

and leave the final choice to him. Such procedure, however, can never be routine. The counselor must exercise judgment and caution in presenting the test results.

Space does not permit presentation of enough of these profiles to illustrate all observed patterns; the above are sufficient to indicate the method. We may suggest that in other cases more difficult problems sometimes arise, as, for example, in the situation where there are no high scores, or where the score on the chosen occupation is low, or where there is a very strong drive toward a profession accompanied by low intelligence. The interest inventory does not solve all these problems, and it would be unreasonable to expect it to do so. It does, however, summarize some relevant facts. The method of graphic analysis presented here seems to furnish a basis for scientific consideration of facts relevant to vocational choice. The method, if studied with sufficient care and persistence, may yield a procedure sufficiently practical to be of value to counselors.

VII. SUMMARY AND CONCLUSIONS

A preliminary study of the vocational interests of two hundred eight tenth-grade students has been made, utilizing the Strong Vocational Interest Blank and evidence from various other sources. Treatment of the interest-test results emphasized the use of an abbreviated scoring in the interests of economy; the scoring has thus been limited to the seven diagnostic scales recommended by Strong,[12] plus the scale appropriate to the chosen occupation of each student, and additional scales for measuring interest maturity, studiousness, and "masculinity." In the treatment of the data, profiles of standard scores have been used in the attempt to study patterns of interests. Preliminary studies have been made of the stated vocational choices of the students, and available evidence concerning intelligence and economic status has been used to supplement the test results.

The analysis of the data justifies the following tentative conclusions:

1. The Strong Vocational Interest Blank yields intelligible results with high-school students when extended and careful analysis precedes the interpretation of scores. It is obvious that this test can not be routinely applied in the tenth grade.

2. A majority of tenth-grade students will state a choice of occupation when asked to do so. It seems reasonable to suppose that these

choices are somewhat unreliable, but the data show that such choices are significantly related to test scores.

3. High ratings on the vocational interest scales, based upon adult norms, are frequently found for high-school students. In about two-thirds of the cases, ratings above *C* are secured for the chosen occupation.

4. In about two-thirds of the cases, high-school students secure low ratings on occupational scales which are arbitrarily selected. This tendency is even more pronounced in the case of students who have not made a choice of occupation.

5. In the study of sex differences, further data are required. We consider that the Strong Vocational Interest Blank for men can provide significant information concerning the interest patterns of girls. The usefulness of the test for this purpose is limited, and the working out of appropriate methods is a problem for research.

6. Children who state an occupational choice in the test situation when asked to do so, are likely to be brighter, younger, more mature, more studious, and less "masculine" than the children who are unable to state any choice. The children who cannot state a choice are likely to secure higher scores on the scales for Engineer and Physicist.

7. The profile study of scores of high-school students shows that many high-school sophomores have well-developed patterns of interests.

8. These findings are essentially similar to those generally known to be found for adults, and they justify the conclusion that the inventory can be profitably used in the high school when suitable normative procedures are worked out.

9. When the interest inventory is used as one factor in a clinical situation in which other available information is also included with proper emphasis, the inventory will probably be found useful in counseling work.

REFERENCES

1. Berman, I. R., Darley, J. G., and Paterson, D. G.: *Vocational Interest Scales.* Bull. Employment Stabilization Res. Inst., University of Minnesota, Vol. III, 1934, pp. 215–245.
2. Carter, H. D.: "Twin Similarities in Occupational Interests." *Jour. Educ. Psychol.*, Vol. XXIII, 1932, pp. 641–655.
3. Carter, H. D., and Strong, E. K., Jr.: "Sex Differences in Occupational Interests of High-School Students." *Personnel Journal*, Vol. XII, 1933, pp. 166–175.

4. Carter, H. D., Pyles, M. K., and E. P. Bretnall: "A Comparative Study of Factors in Vocational Interest Scores of High-School Boys." *Jour. Educ. Psychol.*, Vol. XXVI, 1935, pp. 81–98.
5. Carter, H. D.: *The Measurement of Masculine and Feminine Attitudes by Means of the Interest Inventory.* Unpublished Thesis, on file in University of California Library, 1933.
6. Cowdery, K. M.: "Measurement of Professional Attitudes." *Jour. Personnel Research*, Vol. V, 1926–1927, 131–141.
7. Fryer, Douglas: *The Measurement of Interests.* New York: Holt, 1931, 488 pages.
8. Mackaye, David L.: "The Fixation of Vocational Interest." *Amer. Jour. Sociol.*, Vol. XXXIII, 1927, pp. 353–370.
9. Paterson, D. G., and Darley, J. G.: *Men, Women, and Jobs.* Univ. of Minnesota Press, Minneapolis, Minnesota.
10. Remmers, H. H.: "The Measurement of Interest Differences Between Students of Engineering and Agriculture." *Jour. Applied Psychol.*, Vol. XIII, 1929, pp. 105–119.
11. Segel, D., and Brintle, S. L.: "The Relation of Occupational Interest Scores to Achievement Test Results and College Marks." *Jour. Educ. Res.*, Vol. XXVII, 1934, pp. 442–445.
12. Strong, E. K., Jr.: *Manual for Vocational Interest Blank for Men.* Stanford University Press.
13. Strong, E. K., Jr.: *Manual for Vocational Interest Blank for Women.* Stanford University Press.
14. Strong, E. K., Jr.: "Classification of Occupations by Interests." *Personnel Jour.*, Vol. XII, 1933, pp. 301–313.
15. Strong, E. K., Jr.: "Interest Maturity." *Personnel Jour.*, Vol. XII, 1933, pp. 77–90.
16. Strong, E. K., Jr.: "Aptitudes *versus* Attitudes in Vocational Guidance." *Jour. Applied Psychol.*, Vol. XVIII, 1934, pp. 501–515.
17. Strong, E. K., Jr., and Carter, H. D.: "Efficiency Plus Economy in Scoring an Interest Test." *Jour. Educ. Psychol.*, Vol. XXVI, 1935, pp. 579–586.
18. Strong, E. K., Jr.: "Permanence of Vocational Interests." *Jour. Educ. Psychol.*, Vol. XXV, 1934, pp. 336–344.
19. Strong, E. K., Jr.: "Predictive Value of the Vocational Interest Test." *Jour. Educ. Psychol.*, Vol. XXVI, 1935, pp. 331–349.
20. Thurstone, L. L.: "A Multiple Factor Study of Vocational Interests." *Personnel Jour.*, Vol. X, 1931, pp. 198–205.
21. Vernon, P. E., and Allport, G. W.: "A Test for Personal Values." *Jour. Abn. and Soc. Psychol.*, Vol. III, 1931, 26.
22. Young, C. W., and Estabrooks, G. H.: *Scale for Measuring Studiousness.* Stanford University Press, April 1936.

A STUDY OF CLINICAL PREDICTIONS OF STUDENT SUCCESS OR FAILURE IN PROFESSIONAL TRAINING*

JOHN G. DARLEY

The General College, University of Minnesota

At some point in the total span of education, students must be distributed to curricular offerings that provide for vocational or professional training. While we may grant that job-training is not the sole objective of education, it is still an end of major importance in the minds of most students and parents, particularly at the college level. Increasing college enrollments and greatly expanded curricular offerings complicate this already difficult function of distributing students to curricula that best meet their occupational training needs.

In the high-school years, students are distributed primarily into three broad groups: General academic or college preparatory curricula; commercial courses; and trade or vocational training. The majority is found in the general academic courses, even though less than half of the high-school students go on to college or university training. In Minnesota, there is some evidence to indicate the operation of irrelevant or unimportant factors in the selection of high-school seniors planning to go on to college. Many poor students plan to enter higher academic levels; many good students make no further educational plans.[11] Pre-college distribution of students to higher education tends to be based upon factors extraneous to the abilities and interests basic to college success: Financial considerations; lack of awareness of superior abilities; proximity to institutions of higher learning; and family pressures seem to play overly important parts in defining the calibre of students who present themselves for college admissions. However, once these students have applied, a study by the Minneapolis School Board[14] shows a tendency toward selection of potentially better academic risks by institutions of higher learning.

* The author wishes to acknowledge the assistance of W.P.A. project #6094-4214 in analyzing some of the data for this study. He is also indebted to Dr. M. S. MacLean, Director of the General College, for the latitude permitted to the counseling division in acting on transfer applications and in making comparative analyses of records under two transfer systems. Finally, without the excellent coöperation of the administrators of the receiving colleges, particularly Dr. R. R. Shumway, Assistant Dean of the College of Science, Literature and the Arts, the study would have failed for lack of a testing ground for the clinical predictions of success or failure.

While such a state of affairs is to some extent an indictment of high-school guidance and personnel programs, higher institutions are far from meeting the distributive problem adequately, as evidenced by the high student mortality rate at the college level.

When the high-school senior decides to go on to college, he files an application for admission that, in most cases, calls for a statement of his vocational choice. Here again, family pressures, desires for economic security, romanticism, prestige values, and white-collar complexes operate as irrelevant or illogical bases of choosing occupations. In another study of Minnesota high-school seniors, forty to forty-five per cent of the men consistently chose engineering, law, medicine, scientific research, and higher business executive occupations[12] as their goals. Proponents of the "onward and upward" philosophy may find such ambition praise-worthy, but it is extremely doubtful if forty to forty-five per cent of high-school male graduates have sufficient ability to complete training for these occupations or could find positions after graduation. Society is not geared to absorb or support such a proportion of professional men.

After these inexperienced and somewhat randomly selected high-school students have made their preponderantly illogical choices, the college or university swings into action. Patterns of high-school academic and non-academic credits are evaluated by admissions committees or registrars' offices. The proper amount and kind of high-school credits admit the student to the start of the curriculum of his own choice. But Heaven help the student from the high-school commercial curriculum who later decides to enter college! The red tape, the probationary period, and the obstacles of required credits to be made up are effective deterrents to progress. Traditionally, patterns of high-school credits were assumed to be one adequate basis of predicting subsequent college success. Recent research, fortunately, has raised considerable doubt as to the validity of such high-school patterns, with the result that admissions policies are coming to place greater emphasis on past achievement and measured ability as more valid indices of academic success than strict adherence to pattern of high-school content. The national study sponsored by the Progressive Education Association, in which coöperating colleges have agreed to accept graduates from coöperating high schools without regard to high-school pattern is but one example of experimental approach to this topic. Given adequate personnel work with the students involved, this study should be eminently successful. Kronenberg[6] lists and critically reviews other studies bearing on this topic,

indicating again the decreasing emphasis on patterns of high-school subjects in admissions policies. At Minnesota, students with incomplete high-school records may receive special consideration for admission to the college of their choice on the basis of special entrance test results.

If the student clears the hurdle of general entrance requirements, in addition to the special requirements for the curriculum of his choice, he is next faced with the problem of selecting specific courses from the wide range of available offerings. If he fails to meet entrance requirements, he is set to work at tasks judged to be try-outs for the field of his choice or sent to some other curriculum, the requirements for which his high-school pattern meets. In either event, his study program must be planned. Departmental and staff advisors are on hand for a freshman orientation period prior to the opening of classes; these advisors map out courses in terms of more stated prerequisites, analogous to high-school patterns; or in terms of generalization requirements; or in terms of the student's expressed choice. In the great majority of cases, these advisors know little or nothing about the abilities, interests, specialized needs, or background of the students whose programs they plan. Continuing general advising or departmental advisory systems are usually available for work with students during the school year; in some instances faculty advising is based on adequate diagnosis of the student's characteristics or potentialities. Such instances are so rare, however, that the usual faculty advisory system may be said to operate in the dark as far as valid knowledge of the individual student is concerned.

Other methods of helping students distribute themselves may be briefly sketched. The uniform freshman year is designed to give students a chance to find themselves before making a final decision; orientation or survey courses are set up to integrate fields of knowledge for the student; vocational lectures are given by leaders in various occupations to aid students in making a choice of a life work; variations of the advisory system plan are set up. In all these methods there is again lacking any system of individual diagnosis to determine whether the educational and occupational choices so made are consonant with the interests, abilities, and needs of the student. It is tacitly assumed that the student will choose wisely; such an assumption is made about the student who a few short months ago, as a high-school senior, was confused or incapable of choosing wisely under much the same mass guidance methodology in high school.

Related services such as housing bureaus, speech clinics, health services, how-to-study classes, scholarship or loan funds, social programs, and other forms of individualized attention are frequently provided, depending upon the resources of the institution. These resources are designed to clear up extra-educational problems of potential danger to the student's achievement and development in the curriculum to which he has been distributed.

Assuming now that students are distributed to curricular offerings, they have still other tasks ahead of them: To complete each segment of the prescribed curriculum in proper order; to fill in the proper number of electives; and above all to keep their work up to the standards required by the institution. Failure in these regards means exclusion or suspension from the curriculum of their choice.

We may digress for a moment to consider this question of curricular content and academic standards. We may ask, for example, whether the content of a curriculum is closely related to, irrelevant for, or adequate to the life and job needs that the student will face after leaving school. Has content been determined by an approximate form of job and worker analysis? To this, the professional curricula may answer, "yes." But where the job situation is not so accessible for analysis, or where the life situations to be faced are not clearly visualized, how is content determined? By logic, armchair or otherwise; by tradition; by expert judgment; by disclaiming responsibility for "practical" training; by a retreat to the "classics"; or by envisaging the student in the likeness of the faculty? This problem is the topic of another article, where a methodology of deriving a clear picture of the *students* is described as a basic factor in curricular-planning;[1] it is mentioned here because of its bearing upon distribution of students, since adequate distribution involves meeting *student* needs and building on *student* abilities and interests in the classroom situation. In this respect, a thoroughly adequate case could be established for defining *effective* education as that education which has the highest retention value in subsequent life adjustments made by the student.* Such a definition is a generalized application of Ebbinghaus' experiments on memory, and the curves of retention or forgetting, wherein learned material kept in use can be shown to have greater retention value. If the test of retention were applied to much of the existing

---

* This concept is being further elaborated in forthcoming publications by J. G. and K. McC. Darley.

academic content, there might be a valid question as to the usable residue accruing from our present selection of material for classroom presentation.

Continuing the digression, how are standards set? Are they equal to, higher than, or lower than the standards the individual must meet in life and job situations? Is the grading system by which standards are maintained either accurate or meaningful? A grade of *C* varies, in the amount of achievement it represents, from one institution to another, among colleges in one institution, among instructors in one department, and finally the instructor himself is a changeable measuring instrument. Are not "standards," then, the mere simulation of objectivity, if not actually false measures of achievement? It has been demonstrated in one situation that the ability level of a college moved upward over a period of years, yet the grading system kept one jump ahead, resulting in the assignment of failing grades to approximately the same proportion of students in two divergent ability groups.[10] It has also been demonstrated that, by maintaining the same measuring instrument and the same grading system, failures can be seen to vary in relation to the ability and motivation range of changing classroom groups.[9] Valid and meaningful content, taught by adequate classroom techniques to students whose needs such content meets, with achievement measured on a stable and accurate examination and graded in terms of a basic norm group, will yield standards that can be relied upon. But when students are assigned failing grades in a frame of reference dissociated from their abilities, "standards" are synonymous with rubber yardsticks. These points are raised only because they relate to the student's continuation in the curriculum to which he has been distributed.

Even with the use of one or more of these distributive methods, and within the framework of predetermined content and academic standards, student mortality at the college level continues at the rate of fifty to seventy-five per cent. The distributive system and the method of evaluating the achievement of the distributed students may both contribute some share of this mortality rate. A survival rate of one or two out of four entrants is defensible at all only if education adheres to the single obligation of intensive training for technical and professional leadership. If, however, education is to be considered an instrument to provide for more efficient human adjustment on many social, personal, and vocational fronts, the mortality rate becomes clearly less defensible. We are not here crediting all student mortality

to deficient methods of distribution of students, or to vagaries of grading systems. But, knowing from clinical experience the illogical bases of students' educational and vocational choices, and knowing from research studies the inaccuracies of grading methods, the limited predictive value of high-school patterns and prerequisites, we feel little hesitancy in crediting to the distributive function a definite, if undetermined, share of student mortality.

The data to be presented represent a study of several aspects and interrelations of the distributive function as it affects the student population in the General College of the University of Minnesota. Distribution as a definite phase of student personnel work[13] may be defined as the selection, with and for the student, of the curriculum which best fits his individual needs and in which his chances of successful competition are maximized. This represents, in part, an adaptation of the concept of occupational ability profiles and families of occupations[8] to the educational system. It is also the same basic definition, with slightly different implementation, as that given by Koos and Kefauver.[5] It depends for its success upon *research* measurements of student and occupational groups, with such measurements *clinically* interpreted for the individual student. The opportunity to organize such a program was afforded by a General Education Board grant to the General College in the summer of 1935; under the grant, two full-time personnel workers were set the objective of studying the students. The personnel program was set up in the first year with the major emphasis on service to students, to permit the collection of necessary clinical data on the individual. One major aspect of this service program involved making recommendations to the administration on applicants for transfer to professional curricula in the other colleges and divisions of the University. It is to be emphasized at the outset that this activity was carried on under pressure from students, parents, and other sources, such that clear-cut and definitive research results were doubtful. The same pressures had operated more directly on the administrators in the college, who had cleared transfer applications prior to the establishment of the personnel program, using where possible personnel services existing elsewhere on the campus. Starting in the fall of 1935, the personnel system introduced a standard testing program for entering freshmen and also persuaded a large sample of returning students to coöperate in the same testing program. These tests, lasting four to five hours, provided the basic measurement data later

to be used clinically in counseling individual students. A more systematic record-keeping procedure was also introduced to permit later research analysis of the student personnel files.

As the problem of selection of students for transfer to professional curricula became increasingly pressing, attempts were made to introduce refinements and formalization of the recommending function. Differential recommendations, based on differential predictions of success or failure, were set up in case notes or clearance forms. To provide material for analysis, transfers in the academic year 1934–1935 were traced and their transfer recommendations and predictions reinterpreted or recast in terms of the categories growing out of the personnel work in 1935–1936. It was possible to compare the success of students transferred under a primarily administrative system with students transferred under a personnel system having facilities for placing greater emphasis on clinical diagnosis of student potentialities.

Table I presents data on the sources of student supply in the General College for the peak year of the college's enrollment (1935–1936); and the proportions of each source of student supply transferred by the administrative and by the personnel systems, systems *A* and *P* respectively, during the seven academic terms covered by the study—fall quarter, 1934–1935 through the fall quarter, 1936–1937. The break-down in columns 1 and 4 has been taken from an unpublished Hollerith analysis of case records of students in residence some time during 1935–1936. Figures in columns 2, 3, 5, and 6 relate specifically to students transferring from the General College in any one of the seven quarters under analysis, regardless of time of residence in the General College.

Table I first indicates the high degree of selection found in the General College to begin with. The students have already been distributed in terms of distributive policies operative at Minnesota. These distributive policies utilize nearly all the usual methods described earlier, and, in addition, represent a gratifying emphasis on the results of personnel research, as well as the clinical recommendations of trained personnel workers for individual cases. Summing categories 4 and 6, one finds the effect of distribution in terms of high-school pattern; summing categories 2, 3, and 5, one finds the group whose previous achievement and/(or) measured ability make them questionable academic risks at the time of admission to university work; category 7 includes students who tried and failed a curriculum of their choice; categories 1 and 9 include students to whom the General College

TABLE I.—REASONS FOR REGISTRATION IN THE GENERAL COLLEGE FOR ALL 1935–1936 STUDENTS, AND FOR STUDENTS TRANSFERRED FROM THE GENERAL COLLEGE UNDER TWO TRANSFER SYSTEMS DURING SEVEN ACADEMIC QUARTERS

Reasons for general college registration	1* All men 1935–1936		2 Transfers System *A* men		3 Transfers System *P* men		4* All women 1935–1936		5 Transfers System *A* women		6 Transfers System *P* women	
	N	Per cent	*N*	Per cent	*N*	Per cent	*N*	Per cent	*N*	Per cent	*N*	Per cent
1. College of choice	115	14	15	8	11	9	85	18	2	3	5	8
2. Advised against Arts College entrance (college ability rating in range of doubtful success†)	144	18	40	23	15	12	68	14	11	14	8	12
3. Refused for Arts College entrance (college ability rating in failure range†)	108	14	15	8	10	8	27	6	1	1	1	2
4. Refused by college of choice for incomplete high-school record (all colleges except institute of Technology)	123	15	36	21	39	33	80	17	28	37	20	31
5. Refused by Institute of Technology for poor high-school record	16	2	4	2	4	3	0	...	0	...	0	...
6. Refused by Institute of Technology for incomplete high-school record	31	4	12	7	7	6	1	...	0	...	0	...
7. Transferred to the General College after poor record in other colleges at the University	203	25	38	21	33	27	149	32	29	38	26	40
8. Transferred from outside colleges or universities	24	3	8	5	0	...	13	3	2	3	1	1
9. Transferred to the General College with clear record in another college at the University	26	3	4	2	2	2	24	5	0	...	3	5
10. Assigned, not otherwise specified	15	2	5	3	0	...	22	5	3	4	1	1
11. Totals	805	100	177	100	121	100	469	100	76	100	65	100

* Based on typical figures for 1935–1936 and includes one hundred fifty-seven formal refusals of transfer requests, as well as recommendations for transfer in the same year.

† College Ability Rating is the average of percentile rank in high-school graduating class and percentile rank in University entrance test.

program seemed sufficiently appealing to warrant its choice. The net result of prior distribution of the kind indicated in Table I is morale problems, student and family pressures, and indeterminate effects on motivation, in spite of the fact that the General College curriculum much more closely approximates the needs of the group than did the traditional curricula which previous generations of similar student groups had tried and preponderantly quit within two years, prior to the establishment of the General College. Category 4 is somewhat misleading in that, under existing admission policies, at the University of Minnesota, many of these students would have fallen in categories 2 and 3, had their high-school pattern been intact.

It was the difficult and exacting job of both the administrative and personnel systems to make a further and presumably more accurate distribution from this already highly selected student body. Student wants were phrased in curricular choices usually beyond their capacities; student needs were presumably being met in the curricular research leading to revisions of the General College curriculum itself. The conflict between wants and needs yielded considerable individual pressure on occasion. Such a conflict is not unique in the General College at the University of Minnesota. Eells[2] has presented an excellent and extensive study of further college intentions of almost seven thousand junior-college students in California. All of these students expected to go on for further training, but only twenty-eight per cent actually did so and only seventeen per cent actually graduated. About half of those continuing in higher education graduated, and this proportion of graduates varied inversely with the difficulty of the professional curricula selected. Harbeson[4] points out the fallacies, errors, and difficulties encountered in methods of segregating high-school or junior-college students into perfect dichotomy of terminal and professional degree candidates. Junior colleges are actually terminal educational units for the greater majority of students, even though the students themselves tend to seek further education and tend to fail when they get to the curricula of their desires. At Minnesota, the problem is intensified, the proportions of professional training intentions is greater, and the proportion of completed intentions (graduation) is smaller, because the General College operates in physical conjunction with other units of the University and the General College students are drawn from populations who previously failed in professional training in these other units.

The differing proportions of students transferred from each source of student supply by the two transfer systems are also presented in Table I. Both systems transferred about the same proportion of students who had previously failed presumably to curricula in which they could succeed; System *A* transferred a greater proportion of men students originally judged to be poor academic risks; System *P* transferred a greater proportion of men students with deficient high-school patterns. These discrepancies are due in part to the greater amount of individual data collected under System *P*.

How successful were the two systems in their distribution of students to professional training? The measure of success in each case was an honor point ratio* exactly equivalent to a letter grade of *C*, except for the Institute of Technology transfers, where the passing grade is known to be *D*. This is a rigid measure of success, since many colleges do not demand exactly a *C* grade for continued residence. However, since no arbitrary grade would be an adequate composite of the continuation requirements for all colleges, the grade of *C* is no better or no worse as an arbitrary measure of success than any other grade. The results are subject, in addition, to the shifting value of the *C* grade itself, in various colleges, but this uncontrolled factor must remain uncontrolled under present marking systems. Its potential effect was handed only by differential transfer recommendations on differential predictions of success.

Predictions were phrased or interpreted finally in four categories: unconditional success; doubtful prediction; immediate or eventual failure; and conditional prediction. There was also a category of "no prediction made or recorded" to cover cases where administrative action alone effected transfer. A doubtful prediction covers cases in which interpretation of existing data is questionable and where an actual try-out provides the only answer; such cases might look like bad risks but not bad enough to predict failure. A conditional

* In deriving an honor point ratio, four points were assigned for each hour of grade *A*, three points for each hour of grade *B*, two points for each hour of grade *C*, one point for each hour of grade *D* or *E*, and 0 points for each hour of grade *F*. Hours cancelled without grade and hours of incomplete grades were not used in the calculations. Grade records through the fall quarter of the year 1936–1937 were collected with the coöperation of the Registrar's Office. Thus, for System *A*, students transferred in the fall quarter of 1934–1935 might conceivably have reported up to seven quarters of professional training; while for System *P*, the follow-up of the last group, transferred at the beginning of the fall quarter, 1936–1937, would cover only one quarter of professional training.

prediction indicated success *if* certain extra-educational or educational problems were cleared up before or during professional training. These might be family or emotional or financial problems, in addition to choice of curricula, study habits, or other specific academic handicaps. Depending upon existing relations with the receiving college,

TABLE II.—SUMMARY OF STUDENTS (ALL MEN) TRANSFERRED FROM THE GENERAL COLLEGE DURING SEVEN QUARTERS UNDER TWO SYSTEMS OF CLEARING TRANSFER APPLICATIONS

Group	Form of prediction and result	System *A*				System *P*			
		N	Per cent	*N*	Per cent	*N*	Per cent	*N*	Per cent
1	Success predicted	..	..	98	57	..	..	45	39
2	Failure predicted	..	..	2	1	..	..	22	19
3	Achievement as predicted	49	49	...	...	43	64		
4	Success; doubtful or conditional prediction	4	25	...	...	9	28		
5	Failure; doubtful or conditional prediction	12	75	...	...	23	72		
6	Total doubtful or conditional predictions	..	..	16	9	..	..	32	28
7	Success; without prediction	14	24	...	...	8	50		
8	Failure; without prediction	43	76	...	...	8	50		
9	Total transferred without prediction	..	..	57	33	..	..	16	14
10	Total transfers	..	..	173*	100	..	..	115*	100
11	Total successes, regardless of prediction	..	..	65	...	..	..	50	
12	Per cent successes, regardless of prediction	..	..	...	38	..	..	...	43

* Excluding ten students leaving University immediately after transfer.
Critical Ratios: Achievement as Predicted: Systems *P vs. A.*: 1.948.
Per cent Successes, Regardless of Prediction: Systems *P vs. A*: .847.

doubtful predictions might actually be predictions of failure in the case notes, but in order not to militate against the student's chance of acceptance on transfer, the formal prediction would hold out some possibility of successful work. Altogether, the men and women who transferred went to thirteen other colleges or divisions within the University of Minnesota. Thus the administrators or personnel workers in the college had to be somewhat familiar with the prevailing

requirements and standards of these units. The College of Science, Literature and Arts, the College of Education, and the College of Agriculture, Forestry and Home Economics, in that order, accepted the three largest proportions of General College students.

TABLE III.—SUMMARY OF STUDENTS (ALL WOMEN) TRANSFERRED FROM THE GENERAL COLLEGE DURING SEVEN QUARTERS UNDER TWO SYSTEMS OF CLEARING TRANSFER APPLICATIONS

Group	Form of prediction and result	System *A*				System *P*			
		N	Per cent	*N*	Per cent	*N*	Per cent	*N*	Per cent
1	Success predicted	..	..	49	64	..	..	30	49
2	Failure predicted	..	..	0	0	..	..	11	18
3	Achievement as predicted	25	51	..	...	27	66		
4	Success; doubtful or conditional prediction	2	22	..	...	1	8		
5	Failure; doubtful or conditional prediction	7	78	..	...	12	92		
6	Total doubtful or conditional predictions	..	..	9	12	..	..	13	21
7	Success; without prediction	6	33	..	...	2	29		
8	Failure; without prediction	12	67	..	...	5	71		
9	Total transferred without prediction	..	..	18	24	..	..	7	12
10	Total transfers	..	..	76*	100	..	..	61*	100
11	Total successes, regardless of prediction	..	..	31	...	..	..	28	
12	Per cent successes, regardless of prediction	..	..	..	41	..	..	..	46

* Excluding four students leaving University immediately after transfer.
Critical Ratios: Achievement as Predicted: Systems *P vs. A.*: 1.453.
Per cent Successes, Regardless of Prediction: Systems *P vs. A.*: .967.

Tables II and III, for men and women respectively, show: The proportions of students transferred under the two systems in each prediction category; the proportion of students definitely conforming to the two prediction categories of success or failure (*i.e.*—proportion of correct predictions); the proportion of students in the categories of doubtful or conditional prediction who failed or succeeded; and finally, regardless of prediction, the proportion of transfers succeeding under each transfer system.

Certain differences are immediately observable under the two systems. System *P* was less willing to predict unconditional success, and more inclined to interpret clinical and measurement evidence as indicative of failure in professional training than was true of System *A*. System *P* had a smaller proportion of students in the "no prediction" category than did System *A*. Of the students in this "no prediction" category for System *A*, over seventy per cent of men and women failed after being transferred, as compared to about fifty-seven per cent of failures under System *P* in the same "no prediction" category. In the "doubtful or conditional prediction" category, the greater majority of students failed under both systems.

The two percentages for which critical ratios were calculated for each sex are of particular interest as measures of the efficiency of the two systems. These percentages and ratios appear in Tables II and III. In the category of "achievement as predicted," System *A* shows a smaller percentage of accurate prediction for both men and women. When the two sexes are combined, in order to compensate for the small samples, the critical ratio for per cent accuracy of prediction becomes 2.46, indicating that the difference between the two systems would recur in at least ninety-nine repeated analyses out of one hundred. Combining the sexes to determine per cent of successes, regardless of prediction, the critical ratio becomes 1.43. In terms of a scheme of differential predictions, the critical ratio covering per cent accuracy of prediction is the more definitive measure of efficiency of the two systems.

One further bit of evidence may be adduced from the footnote of Table I. It was there stated that one hundred fifty-seven students were formally recommended, preponderantly under System *P*, for refusal of transfer applications. In these cases the evidence was such that failure was uniformly predicted and the pressures were such that these refusals closed the case at that time. Comparable data for refusals under System *A* were not available for this study. Since failure is uniformly more amenable to prediction than is success, these refusals may, qualitatively at least, be listed as correct predictions under System *P*; especially since the average ability level of the refusal group was found in other research to be not only lower than the transfer group, but also lower than all other students not applying for transfer.

Fourteen men and women, under both systems, were transferred with the conditional predictions defined earlier. Of these, two suc-

ceeded, one cancelled from the University, and the remaining eleven fell below the *C* grade. In ten of these cases, extremely severe emotional or home or health conflicts existed. Two other students were undiagnosed non-achievers and the remaining two were working under marked financial handicaps. No generalizations can be drawn from such a small group; nor were these the only students laboring under extra-educational handicaps. Such handicaps were confined to no one prediction category; in these few cases, however, they conditioned the predictions, with the results indicated. It is conceivable that the "conditional" prediction category may decrease under more refined techniques of analysis, just as the "doubtful" category may decrease.

Thus the trend of evidence consistently indicates the increased efficiency of distribution of students under a system emphasizing student personnel methods of distribution as defined earlier. Even though the population under analysis was an already segregated group, as defined by local academic standards and admissions policies, it was possible to identify an appreciable number of students who could succeed in professional training, in spite of their earlier classification as bad academic risks.

There are other problems on which these data have limited bearing. The first is the question of rehabilitating students who fail in one curriculum. Table I showed the proportion of students in the General College because of poor records elsewhere in the University; Table I also showed the proportion of such students transferred under the two systems. Tables IV and V, for men and women respectively, represent a further analysis of these "rehabilitation" cases. In these tables, it appears again that the personnel system, in terms of *C* grades as the criterion of success, rehabilitates students more successfully. While the numbers of cases are too small to yield completely reliable percentages, it is to be noted that the percentages of correct predictions and of total successes regardless of predictions follow very closely the comparable figures for the two systems in Tables II and III. These rehabilitation cases represent two broad groups of students: Those choosing curricula not consonant with their abilities, interests, and backgrounds; those in appropriate curricula who developed other educational or extra-educational problems. In either case, special diagnosis and treatment will yield more accurate predictions of success in subsequent professional training than will a period of General

College residence as a "breathing spell," without simultaneous clinical study of the individual.

A departure from the present data leads to certain interesting generalizations. By a subtractive process (not presented here), the data in Table I yield the following groupings:

1. Two hundred three men and one hundred forty-nine women in the General College after failing other curricula and staying in residence in 1935–1936.

TABLE IV.—SUMMARY OF STUDENTS TRANSFERRED, AFTER FAILING IN ONE PROFESSIONAL CURRICULUM, UNDER TWO TRANSFER SYSTEMS (MEN)

	System *A*	System *P*
1. Returned to same curriculum, with success predicted	8	2
2. Returned to same curriculum, with failure predicted	3	5
3. Returned to different curriculum, with success predicted	13	8
4. Returned to different curriculum, with failure predicted	0	5
5. Achievement according to prediction	11 (46 per cent)	15 (75 per cent)
6. Returned with doubtful or conditional prediction	3	8
7. Returned with no prediction	7	1
8. Total returned	34	29
9. Total successes, regardless of prediction	10	11
10. Per cent successes, regardless of prediction	29 per cent	38 per cent

2. Fifty-three men and thirty-four women in the General College after failing other curricula and transferred to professional curricula by either System *P* or System *A* in 1935–1936.

3. Approximately thirty-one of such transfers as defined in item (2) succeeding in their second attempt at professional training.

From these data we may arrive at a *minimum* percentage of mortality due to wrong curricular choice or wrong adjustment to an adequate curricular choice. Of the three hundred fifty-two men and women in the General College in 1935–1936 after failing work elsewhere, approximately eighty-seven earned or were given a second chance in this period. Approximately thirty-one succeeded, which is about nine per cent of three hundred fifty-two. This may stand as

the minimum share of failing students with possibilities for success if given another trial. To arrive at the maximum percentage of rehabilitation possibilities, the following four unknown quantities must be identified:

1. Number of marginal students (barely passing) in present curricula who would be desirable students in different curricula within the University.

TABLE V.—SUMMARY OF STUDENTS TRANSFERRED, AFTER FAILING IN ONE PROFESSIONAL CURRICULUM, UNDER TWO TRANSFER SYSTEMS (WOMEN)

	System *A*	System *P*
1. Returned to same curriculum, with success predicted	6	0
2. Returned to same curriculum, with failure predicted	0	2
3. Returned to different curriculum, with success predicted	12	12
4. Returned to different curriculum, with failure predicted	1	3
5. Achievement according to prediction	7 (37 per cent)	11 (65 per cent)
6. Returned with doubtful or conditional prediction	3	5
7. Returned with no prediction	8	2
8. Total returned	30	24
9. Total successes, regardless of prediction	10	11
10. Per cent successes, regardless of prediction	33 per cent	46 per cent

2. Number of students who fail, but for whom failure is symptomatic of adjustment problems that might be cleared up under more careful clinical treatment.

3. Number of students who withdraw to try non-University vocational training in which they have greater chances of success.

4. Number of students failing in professional curricula who continue for the non-professional general education degree at the University.

If the proportions in these four groups were known, the maximum share of student mortality attributable to invalid curricular choices might be gauged by additive or subtractive processes. In such an analysis of college mortality, no one cause alone would cover proportions of failures, but primary conditions or reasons for leaving college might be identified. In Item 2 above would be found the major social, or emotional, or personal problems basic to academic deficiency. To

financial, health, and specific educational problems might also be attributed a share of student mortality. If a standard measurement and clinical program could be applied to an adequate sample of students leaving school, many facts about student mortality could be identified. While this is an interesting point of speculation, it goes well beyond the present data.

Another problem lies in possible errors of measurement when tests are used in segregating student groups. As was mentioned earlier, admissions to the College of Science, Literature and the Arts are based not only upon intact high-school patterns, but also upon measured evidence, in high-school achievement and in entrance tests, of potential scholarship. Williamson[10] has presented evidence on increases in Arts College ability levels at Minnesota under this admissions policy. Categories 2 and 3 in Table I show the incidence of students whom the Arts College tends to exclude as doubtful or bad academic risks. To determine possible errors of measurement in these predictions, the records were traced for students transferred to the Arts College, or related work, from this defined low-ability group. System *A* transferred thirty-one men and women without any retesting and twenty men and women with retests that made slight changes in the original measurements. System *P* transferred twenty men and women without retests and four men and women with retests. The absence of a retest does not mean that retesting would not have been advisable. Twenty students under System *A* and seven under System *P* succeeded after transfer, regardless of predictions. By making approximations from figures in the Registrar's enrollment report for the biennium 1934–1936,[7] the figure of one thousand nine hundred sixty-six separate student units was judged to be the best estimate of enrollment in the two academic years 1934–1935 and 1935–1936. Taking as typical the per cent of students assigned to the General College in 1935–1936 in Categories 2 and 3 as bad or doubtful academic risks for Arts College work (twenty-eight per cent), approximately five hundred fifty of the one thousand nine hundred sixty-six separate students should be so classified over the two-year period. Of these five hundred fifty, seventy-five students earned or were given transfer by both Systems *A* and *P* in 1934–1935 or 1935–1936. Twenty-seven succeeded, which represents five per cent of the original five hundred fifty so assigned to the General College. This five per cent can be considered the minimum amount of error in the prediction of *failure* in Arts College work under existing admissions policies. A few students not transferred in

the original five hundred fifty might have succeeded if they had been given the opportunity. Such cases would increase the margin of error in the prediction of failure, but pragmatically the margin of error may be set at the five per cent determined above. The margin of error in predicting success, however, would be much higher, since it is demonstrably easier to predict failure than it is to predict success, in situations of this kind. In a scatter-diagram with academic ability plotted on the ordinate and academic achievement on the abscissa, the upper left-hand quadrant contains the non-achieving students, who consistently out-number the over-achievers falling in the lower right-hand quadrant.

Students originally excluded from the College of Science, Literature and the Arts because of incomplete high-school records, but later transferred to that college or its related professional work represent another special group for analysis. In their cases is found the application of the theory that high-school pattern is a valid index of potential college achievement. In the total group of students transferred from the General College, this incomplete high-school-record group stood highest, on the average, in high-school rank, entrance test, and subsequent Arts College grade records. When compared with samples of students admitted directly to the Arts College with complete high-school records, however, they were lower in the high-school achievement and entrance test indices, which supports the statement made earlier in this article that even with complete high-school records they might have been advised against trying Arts College pre-professional training as doubtful academic risks.

Table VI summarizes the statistical analysis of this group of transfers, and also includes comparable data adapted from Tables II and III in Williamson's[10] article.

When high-school rank, entrance test, and General College grade record were averaged as three equal indices of Arts College success, the correlation with Arts College grades for the eighty transfers was increased to only .36, an increase so slight that General College grades seem to have added little to the predictions that could have been made from entrance records alone for these cases for their later Arts College work. The comparable correlations in Table VI, adjusted for size of General College samples and interpreted in terms of standard errors of the various correlations, tend to indicate that complete or incomplete high-school patterns are less predictive of Arts College achievement than are measures of high-school achievement and meas-

ures of academic ability. Granted that General College transfers may not have achieved as well as an unmatched group of direct entrants, the achievement differential between the groups is more properly attributed to ability and high-school achievement differentials than to differences in formal high-school patterns.

Just as the validity of high-school pattern may be questioned in predicting college achievement, so may the validity of college pre-professional requirements be questioned as predictors of professional college training. In this regard, Eurich[3] found no critical differences

TABLE VI.—MEANS, SIGMAS, AND CORRELATION WITH ARTS COLLEGE SCHOLARSHIP OF DEFINED STUDENT GROUPS

	College aptitude test				High-school rank			
	N	$\bar{X}$	σ	*r*	*N*	$\bar{X}$	σ	*r*
General college transfers with incomplete high-school records	80	43.3	26.8	.15	67	62.2	26.5	.29
Direct arts college entrants with complete high-school records	827*	60.0	....	.25	827	73.0	22.1	.53

* Adapted from Williamson's 1935 entrants, *op. cit.*

in professional training grade records between seventy-seven pairs of students, matched for ability, with one in each pair pre-professionally trained in the Arts College and the other pre-professionally trained in the General College. Individual differences in basic ability were more valid indices of later achievement than locus or pattern of pre-professional training. If comparable data from the Chicago College plan or the University of Florida General College plan bear out this finding, the validity of pre-professional requirements in specific instances will be open to still greater question.

SUMMARY

1. Traditional practices of distributing students to curricular resources are discussed, with some question as to their effectiveness in the light of continuing high student mortality.

2. Personnel techniques are suggested as aids in distributing students.

3. Two procedures used in transferring General College students at the University of Minnesota are evaluated to indicate the greater

accuracy of the personnel program in identifying success or failure, even within a population as highly selected, both by University admissions policies and by University personnel research, as the General College group.

4. The personnel program tended also to be more effective in rehabilitating students who failed once in other curricula.

5. "Errors of measurement" in the prediction of failure are discussed in the light of transfer records of students originally judged to be poor or doubtful academic risks.

6. Patterns of high-school subjects are shown to be less valid as indices of college achievement than high-school achievement and basic measured ability.

BIBLIOGRAPHY

1. Darley, J. G. and Darley, K. McC.: "The Keystone of Curricular Planning." *Journal of Higher Education*, Vol. VIII, No. 1, January, 1937.
2. Eells, W. C.: "Intentions of Junior College Students." *The Junior College Journal*, Vol. VII, No. 1, October, 1936.
3. Eurich, A. C.: "A Study of Subsequent Academic Achievement of General College Transfer Students." *The Effective General College Curriculum.* University of Minnesota Press, 1937, pp. 303–307.
4. Harbeson, J. W.: *Classifying Junior College Students.* Ph.D. Thesis, School of Education, University of Southern California, 1931.
5. Koos, L. V. and Kefauver, G. N.: *Guidance in Secondary Schools.* The Macmillan Company, New York, 1932.
6. Kronenberg, H.: "Validity of Curriculum Requirements for Admission to the General College of the University of Minnesota." *Studies in Articulation*, Committee on Educational Research, University of Minnesota, 1937.
7. *Report of the President for the Years* 1934–1936. University of Minnesota Bulletin, Vol. XXXIX, No. 61, December, 1936, p. 145.
8. Trabue, M. R.: "Functional Classification of Occupations." *Occupations*, Vol. XV, No. 2, 1936, pp. 127–131.
9. Williamson, E. G.: "A College Class in Occupational Information." *School Review*, Vol. XLV, No. 2, February, 1937.
10. Williamson, E. G.: "The Decreasing Accuracy of Scholastic Prediction." *Journal of Educational Psychology*, Vol. XXVIII, No. 1, January, 1937.
11. Williamson, E. G. and Darley, J. G.: "Matching Abilities to Jobs." *The Personnel Journal*, Vol. XIII, No. 6, April, 1935.
12. Williamson, E. G. and Darley, J. G.: "Trends in the Occupational Choices of High School Seniors." *Journal of Applied Psychology*, Vol. XIX, No. 4, August, 1935.
13. Williamson, E. G. and Darley, J. G.: *Student Personnel Work: An Outline of Clinical Procedures.* McGraw-Hill Book Company, New York, 1937.
14. *When Pupils Leave School.* Division of Instruction, Minneapolis Public Schools, October, 1937.

THE ESTIMATION OF SPECIFIC AND BI-FACTORS

GODFREY H. THOMSON

University of Edinburgh

I. INTRODUCTION

The present note has two objects. One is to give a simplified method of calculating the correct estimation of a specific in a hierarchical team, or an estimation of a second group factor in any team which can be reduced to hierarchical form by pooling.[1] The other object is to emphasize the fact that if estimates of several factors from the same battery are to be consistent with one another, each factor must be estimated from all the tests of the battery, *including those tests in which that factor does not appear.* This may seem paradoxical at first sight, but on second thoughts will be seen to be obvious. The error of estimating a factor only from tests in which it is present is seen in its extreme form in estimating a specific factor, and I shall consider that case first, as it throws a clear light on the nature of the mistake.

II. ESTIMATING SPECIFIC FACTORS

In the Appendix to *The Abilities of Man,* at page xviii, Spearman gives the estimate of a person's specific ability as

$$\hat{s}_i = r_{is} z_i \tag{1}$$

where z_i is that person's score in the one test containing the specific, and r_{is} is the correlation of that test with its specific. In the 1932 edition of his book, however, Spearman adds a sentence which makes it quite clear that this single-test method is inferior to the method of first estimating g and then deducing the estimate of s so as to be consistent with the estimate of g, as in equation (6) below.

The single-test method, however, is repeated in Holzinger's *Statistical Résumé of the Spearman Two-Factor Theory;* and in *Preliminary Report* 8 on the Spearman-Holzinger Unitary Trait Study,

[1] I should for completeness add that by estimates throughout this paper are meant regression estimates. Other forms of estimation will be more accurate for special purposes (see Bartlett, M. S.: *Brit. Journ. Psychol.*, Vol. XXVI, 1935, p. 199 and Vol. XXVIII, 1937, p. 97) but the regression estimates give the smallest error variance *over the population of persons* whose factors are estimated

Holzinger and his colleagues explicitly state that they estimate factors only from tests containing them, and would estimate a specific factor from the single test containing it. It is this which there causes their methods A and B to differ so much from what they call "Method C (Thomson's Method)," which, therefore, appears at a disadvantage although it is really the correct method, and moreover need not require all the calculations they there associate with it.

In their *Student Manual of Factor Analysis*, which is, I think, later than *Report* 8, they appear, however, to have come around almost entirely to the view here expressed, though some doubt is left in the reader's mind by lines seventeen and eighteen of their page forty in that *Manual*. They have not, however, as far as I am aware, anywhere stated that their contrast of Method C with other methods in *Report* 8, to the apparent disadvantage of Method C, was based on an error.

That there is a mistake in the simple method of equation (1) is best seen from the fact that that equation puts the men who have taken the test, in the same order of merit for the specific as they are placed in by their raw scores in the test, in spite of the fact that the estimates of their g's are known from the rest of the battery. Take the simple case of two men whose g's as estimated from the battery are very different, but who have the same score in this particular test, a two-factor test. Clearly the man with the high g must be estimated to have a low specific and *vice versa*. Yet equation (1) says they have equal specifics.

The estimate of g by a two-factor or hierarchical battery is made from the regression equation (in its well-known Spearman form for a hierarchy)

$$\hat{g} = \frac{1}{1+S}\sum \frac{l_i}{m_i^2} z_i \qquad (i = 1, 2, \cdots \text{tests}) \tag{2}$$

where $l_i = r_{ig}$ = loading of the ith test with g,

$m_i = r_{is} = \sqrt{(1 - l_i^2)}$,

$$S = \sum \frac{l_i^2}{m_i^2}.$$

Similarly the estimate of s_i should be made by a regression equation, which in this case also can be put into a very simple form. Without loss of generality we can take the test to be the first test in the battery. If R is the matrix of intercorrelations of the tests, with unity in each

diagonal cell, then the regression equation for estimated s_1, which we shall designate $\hat{s}_1$, is

$$\hat{s}_1|R| = m_1(R_{11}z_1 + R_{12}z_2 + R_{13}z_3 + \cdots + R_{1n}z_n) \qquad (4)^1$$

Since the matrix R is hierarchical it can readily be shown that

$$\left.\begin{aligned} |R| &= m_1^2 m_2^2 \cdots m_n^2(1+S) \\ R_{11} &= m_2^2 m_3^2 \cdots m_n^2\left(1+S-\frac{l_1^2}{m_1^2}\right) \\ R_{12} &= -l_1 l_2 m_3^2 \cdots m_n^2 \\ R_{13} &= -m_2^2 l_2 l_3 \cdots m_n^2 \\ &\cdots\cdots\cdots \text{ etc.} \end{aligned}\right\} \qquad (5)$$

whence

$$\begin{aligned} s_1 &= \frac{1}{1+S}\left\{\frac{1+S-\frac{l_1^2}{m_1^2}}{m_1}z_1 - \frac{l_1 l_2}{m_1 m_2^2}z_2 - \frac{l_1 l_3}{m_1 m_3^2}z_3 - \cdots\right\} \\ &= \frac{z_1}{m_1} - \frac{l_1}{m_1}\hat{g} \text{ by eq. (2)} \end{aligned} \qquad (6)$$

If we write equation (6) as

$$z_1 = l_1\hat{g} + m_1\hat{s}_1$$

we see that the estimates $\hat{g}$ and $\hat{s}_1$ satisfy the specification of the test $z_1 = l_1 g + m_1 s_1$, as they should. *All that one has to do, therefore, to estimate a man's specific in a hierarchical team, after estimating his g by*

---

[1] It is of some interest to notice that the estimate of a man's specific in a given test is always proportional to *the difference between his actual score in that test, and his expected score in that test* as estimated from his performances in the other tests in the battery. If he scores exactly what his other performances lead us to expect, he has zero (that is, average) specific ability in that test. We have from equation (4)

$$\hat{s}_1|R| \div m_1 = R_{11}z_1 + R_{12}z_2 + \cdots + R_{1n}z_n$$

The estimate of z_1 from z_2, z_3 . . . and z_n is

$$R_{11}\hat{z}_1 = -R_{12}z_2 - \cdots - R_{1n}z_n$$

whence we have

$$\hat{s}_1 = \frac{m_1 R_{11}}{|R|}(z_1 - \hat{z}_1).$$

This is true for all teams, not only for a hierarchy.

equation (2), *is to subtract* $l_i\hat{g}$ *from his score* z_i *and divide by* m_i. Equations (2) and (6) are merely the forms taken in the special case of a hierarchical team by the general regression equation

$$\hat{f} = M'R^{-1}z \tag{7}$$

where f = the column vector of orthogonal factors,
M = the matrix of loadings, with as many rows as there are tests, and as many columns as there are factors,
$R = MM'$ = the matrix of correlations betweens tests,
z = the column vector of a man's scores.

In the hierarchical case there is no need to calculate R^{-1}; or rather equations (2) and (6) are the simplified but exact equivalent of such a calculation.[1]

We saw above that the estimate for s which is given by equation (6) is consistent with the estimate for g given by equation (2), since these two estimates satisfy the specification equations, which are in the general case

$$z = Mf \tag{8}$$

That this is generally true is seen by substituting the estimates $\hat{f}$ from equation (7) for the true f of equation (8) when we get

$$z = MM'R^{-1}z = RR^{-1}z = z \text{ exactly.}$$

The estimate for a specific by equation (1) on the other hand is inconsistent with the g estimate from the battery, and the two do not satisfy the equation

$$z_i = l_ig + m_is_i$$

A numerical example will make this clear. Let four tests have the correlations shown below, which are perfectly hierarchical.

	1	2	3	4
1	.	.7483	.6928	.5657
2	.7483	.	.6481	.5292
3	.6928	.6481	.	.4899
4	.5657	.5292	.4899	.

[1] Even in the ordinary non-hierarchical case there is no need actually to calculate the reciprocal matrix if it is not otherwise needed (see my note in this *Journal* for January, 1936, near the top of page 48, and below the table on page 51.)

A Spearman analysis gives the loadings

	l	m_1	m_2	m_3	m_4
1	.8944	.4472	.	.	.
2	.8367	.	.5477	.	.
3	.7746	.	.	.6325	.
4	.6325	.	.	.	.7746

If in these four tests a certain man scores

1	2	3	4	
1.3416	1.3844	1.4071	1.4071	standard deviations,

the estimate of his g by Spearman's (*i.e.* the regression) formula is

$$\hat{g} = \frac{1}{1 + S}\sum\frac{l_i z_i}{m_i^2} = \frac{14.0676}{9.5} = 1.48$$

By the rule given above, we then estimate his s_1 as

$$\hat{s}_1 = \frac{1.3416 - .8944 \times 1.48}{.4472} = 3 - 2.96 = .04,$$

whereas by the erroneous rule we should estimate it as

$$\hat{s}_1 = .4472 \times 1.3416 = .60$$

that is, to a man who scores 1.3416 in test 1, and to whom we have already ascribed a g of 1.48, we ascribe by this erroneous method a specific of .60, which would have been sufficient to explain a much higher score than he actually obtained; namely,

$$.8944 \times 1.48 + .4472 \times .60 = 1.5920,$$

whereas the proper estimates give his actual score; namely,

$$.8944 \times 1.48 + .4472 \times .04 = 1.3416.$$

Of course not all estimates of specifics made by the wrong method will be overestimates, all depends on what the team estimate of g is; and the discrepancy will be the less, the nearer a man is to average (*i.e.* zero) in all his scores.

Probably the reason why the erroneous method of estimating specifics has survived so long is that no one particularly wants to estimate a specific. The proper method of doing so, as set out above, has been printed in brief at least twice previously, by Wilson (*Proc.*

Nat. Acad. Sc., 1934, p. 194) in a notation, however, unfamiliar to most psychologists and, therefore, perhaps not widely read, and by the present writer (this *Journal* Jan., 1936, p. 43). The principle involved, however, that every factor must be estimated from the whole battery,[1] is applicable also to the estimation of group factors, and insistence on it is particularly necessary in the case of Holzinger's Bi-factor Method.

III. ESTIMATION IN HOLZINGER'S BI-FACTOR METHOD

In Holzinger's Bi-factor Method each test is analysed into two factors and a specific, the first factor running through all the tests, which fall into groups according to their second factors, giving a "hollow-staircase" pattern thus:

Test	Factor loadings									
a	x	x	.	.	x	.	.	.	.	.
b	x	x	.	.	.	x	.	.	.	.
c	x	.	x	.	.	.	x	.	.	.
d	x	.	x	.	.	.	.	x	.	.
e	x	.	.	x	.	.	.	.	x	.
f	x	.	.	x	.	.	.	.	.	x

As a short cut in estimating the factors Holzinger, then, in the methods B and C in Report 8, pools the tests containing the same second factor. The "reduced pattern" thus arrived at is that of a hierarchical system with two "specifics" in each test, the first "specific" being the group factor of that group, thus:

Tests	Loadings						
$1 = a + b$	l_1	k_1	.	.	n_1	.	.
$2 = c + d$	l_2	.	k_2	.	.	n_2	.
$3 = e + f$	l_3	.	.	k_3	.	.	n_3

where $k^2 + n^2 =$ the m^2 of our former considerations. For such a scheme only slight changes are needed in our former equations. In equations (4) and (5) k_1 replaces m_1, and equation (6) becomes, if

[1] It is, of course, understood that motives of economy of time and labour may cause experimenters to depart from this; but in the cases discussed in this article the proper method is so easy that there is no point in being content with a wrong estimate.

s_1 now means the group factor of a and b,

$$\hat{s}_1 = \frac{1}{1+S}k_1\left\{\frac{1+S-\frac{l_1^2}{m_1^2}}{m_1^2}z_1 - \frac{l_1l_2}{m_1^2m_2^2}z_2 - \frac{l_1l_3}{m_1^2m_3^2}z_3 - \cdots\right\}$$
$$= \frac{k_1}{m_1^2}(z_1 - l_1g).$$

After having estimated the general factor g for a man, therefore, *all we need do to estimate correctly his second group factor, in this reduced bi-factor method, is to subtract $l_i\hat{g}$ from his score in the pooled test containing it, and multiply by k_i/m_i^2 where k_i is the loading of the second factor*, and $m_i^2 = 1 - l_i^2$. This will give the same results as Method C in Holzinger's Report 8, which he calls Thomson's Method; but I would have used it in this form, not in his.

Holzinger's Table 9, and his Table 10 in its third row (Method C) could, therefore, have been obtained by much less arithmetic. For his Table 4, of the intercorrelations produced by the reduced pattern is, of course, perfectly hierarchical—or will be very nearly so if actual values are used—and, therefore, Table 9 becomes unnecessary. The general factor u_1 can be estimated by Spearman's usual formula, (which *is* the regression formula in the case of a hierarchy), and the other factors as explained immediately above. Thus, for example, to estimate the factor d for subject 33 we have

Score in test x_d in (Holzinger's) Table 8		1.94
Estimated u_1 in Table 10,	1.84	
Loading l in Table 3,	0.6827	
	0.6827 × 1.84 =	1.256
Difference		0.684
Loading k from Table 3,	0.5939	
$m^2 = 1 - l^2 =$	0.5339	
	k/m^2 =	1.112
	1.112 × 0.684 =	0.761

The figure in Holzinger's Table 10 is 0.75.

Holzinger's Method B estimates the second group factor as 1.15, from the single pooled test only in which it occurs, and is inconsistent with the estimates in the same Table 10 of the factor u_1 (or g). Before we can test this, however, we must estimate by the two methods the third (specific) factor, as is easily enough done. By our present

method it is only necessary to take n/m^2 times 0.684 instead of k/m^2. n in Table 3 is 0.4257.

$$\frac{0.4257}{0.5339} \times 0.684 = 0.545$$

The method corresponding to Method B (equation 1) gives

$$0.4257 \times 1.94 = 0.826$$

We have, therefore, by the two methods

	B	*C* with simplified calculation	Loadings
Estimated *u*	1.84	1.84	0.6827
Estimated *d*	1.15	0.761	0.5939
Estimated specific	0.826	0.545	0.4257

If we multiply the estimates by the loadings, and sum, we get

by *B*	by *C*
2.291	1.940

the first differing from, the second exactly agreeing with, the actual score 1.94.

THE RELATIONSHIP BETWEEN SCORES ON THE SCHOLASTIC APTITUDE TEST AND COLLEGE GRADES IN MATHEMATICS

M. RICHARD DICKTER

Philadelphia, Pennsylvania

THE SCHOLASTIC APTITUDE TEST

In 1926 the College Entrance Examination Board devised an aptitude test, known as the Scholastic Aptitude Test. It is a test offered to colleges and universities for the purpose of testing all candidates for admission to these institutions. Acceptance, however, is entirely on a voluntary basis, for no institution is obligated to give the test.

The test, from 1930 until 1936, the period covered by this investigation, consisted of two sections; one in which the material is essentially linguistic or verbal in nature and another in which the material is essentially mathematical in nature. The test aims to measure potential ability to do work on the college level along specific lines, rather than to measure ability in general. A primary purpose of the Scholastic Aptitude Test is its use as an aid in guidance and teaching in college; for example, selecting students for admission to the various fields of study, guiding them in their choice of majors, forestalling through supervision the possible failures of doubtful students, diagnosing the mediocre performance or failures of promising students, and sectioning groups for instructional purposes.[1]

The Verbal Section.—The Scholastic Aptitude Test was first given to applicants for college entrance by the College Entrance Examination Board in the year 1926. From 1926 to 1929, inclusive, this test consisted only of a verbal, or non-mathematical, section, the composition of which changed during these years. In 1926 and 1927 the verbal section consisted of nine sub-tests. In 1928 the number of sub-tests was reduced to seven, and in 1929 further reduced to six. From 1930 to 1936, the period covered by this study, the verbal section consisted of three sub-tests—Antonyms, Double definitions, Paragraph reading—totalling two hundred items:

---

[1] Bill, E. G. and Pennypacker, H.: "Two Deans Consider the Scholastic Aptitude Tests," *Independent Education*, Vol. III, October, 1929, p. 17.

The Mathematical Section.—Not long after the inception of the Scholastic Aptitude Test, it became evident that the test was better adapted to selecting students for those courses in which verbal abilities are required, such as history, English, Latin, than to selecting students for the mathematical and scientific courses, where mathematical abilities are essential. It was felt that the verbal section alone was not sufficient, because it measured linguistic ability only and not the ability that is thought of as mathematical and scientific.[1]

The dissatisfaction became sufficiently acute to move the Commission on Scholastic Aptitude Tests of the College Entrance Examination Board to add a second section, in 1930, to the Scholastic Aptitude Test, after experimenting with such a section in 1928 and 1929. This second part was a mathematical section consisting of one hundred items, arranged in the order of difficulty, dealing with arithmetic, algebra, and geometry, and designed better to select students for the mathematical and scientific courses.

THE PROBLEM

The Scholastic Aptitude Test is an instrument furnished by the College Entrance Examination Board and is used by colleges and universities as an aid in selecting students for admission to the various fields of study and as an aid in guidance. Considerable time and effort have been, and are being, spent on it. The question arose at the University of Pennsylvania of the degree of relationship between the scores made by students on the mathematical section of the Scholastic Aptitude Test at the time of their entrance to the University and their later academic success, as judged by their marks, in specific courses in first-year mathematics. This study was undertaken in order to answer this question.

At the same time, the relationship between the scores made by the same students on the verbal section and their marks in the same courses was also determined, thus making it possible, in addition, to compare the relative merits of the two sections as far as first-year mathematics is concerned. These relationships were investigated for the period 1930–1936—the years during which both sections of the Scholastic Aptitude Test were given.

The situation is an individual one in each university, so that results obtained at one university are not necessary directly comparable with similar results obtained at other universities. Yet, the findings at

[1] *Annual Report of the Commission on Scholastic Aptitude Tests*, 1929, p. 38.

the University of Pennsylvania, in the sense that they indicate trends, will be of value to other interested groups outside the University.

THE COURSES STUDIED

Five one-semester mathematics courses, generally studied in the freshman year, were included:

Mathematics 2: Plane Trigonometry. A course for beginners.

Mathematics 5: College Algebra.

Mathematics 6: Analytic Geometry.

Mathematics 7*A*: Elementary Analysis. An introduction to graphs, functions, and the differentiation and integration of simple algebraic functions. This course, together with one other course, is recommended for students who wish to take only six semester credits of mathematics in college.

Mathematics 7*B*: Elementary Analysis, continued. Continuation of Mathematics 7*A*, with introduction of trigonometric, logarithmic, and exponential functions. Elementary Analytic Geometry. Prerequisite: Mathematics 7*A*.

THE DATA

All the data for this investigation were obtained directly from the records of the individual students. Likewise, as far as the University of Pennsylvania is concerned, the data do not represent a sampling—all students who came within the scope of the investigation, that is, all students who had been admitted to the University of Pennsylvania during the years from September, 1930, to February, 1936, and who, up to June, 1936, had taken one or more of the five mathematics courses, were included.

In order to obtain the data, all the undergraduate schools[1] of the University were visited and the records of their former and present students examined. When the record of a student was found who had been admitted during the period between the academic years 1930–1931 and 1935–1936 and who had taken one or more of the five mathematics courses, the name of the student, his school, those of the five courses he had studied, and his final mark in each of these courses were noted. Then the admission slips of these students were located and

[1] The College (for men), The College of Liberal Arts for Women, The School of Fine Arts, The Moore School of Electrical Engineering, The School of Education, The Towne Scientific School (Chemistry, Chemical, Civil, and Mechanical Engineering), The Wharton School of Finance and Commerce.

their scores on the verbal and mathematical sections of the Scholastic Aptitude Test recorded.

In this way, a total of four thousand four hundred twenty-three cases, or marks, were obtained for investigation. Of the four thousand four hundred twenty-three cases, fifty-five per cent were found in The College, twenty-three per cent in The Towne Scientific School, ten per cent in The College of Liberal Arts for Women, six per cent in The Moore School of Electrical Engineering, three per cent in The School of Fine Arts, two per cent in The Wharton School, and one per cent in The School of Education.

Mathematics 5 and Mathematics 7*A* contained the largest number of cases, each over one thousand. The analysis of the data showed that the student composition of each course was distributed about as follows:

1. Over half the students in Mathematics 2 were from The College, with most of the remainder distributed about equally between The College of Liberal Arts for Women and The Towne Scientific School.
2. More than half the students in Mathematics 5 were from the two engineering schools, with most of the remainder from The College and The College of Liberal Arts for Women.
3. Most of the students in Mathematics 6 were from the engineering schools, with the remainder in the main from The College.
4. Practically all the students in Mathematics 7*A* and 7*B* were from The College and The College of Liberal Arts for Women.

STATISTICAL TREATMENT

The relationship between test scores and course marks was determined by calculating the coefficients of correlation between the variables involved by means of the Pearson product-moment method. Coefficients of correlation were computed between the scores on the verbal and mathematical sections of students admitted to the University of Pennsylvania in each of the six academic years covered in this investigation, 1930–1931 to 1935–1936, inclusive, and the marks they later made in the courses under consideration. Ten such correlations were made for each year—five between verbal test scores and marks in each of the courses, and five between the mathematical test scores and marks in each of the same courses (verbal test scores and mathematical test scores refer, respectively, to the scores on the verbal and mathematical sections of the Scholastic Aptitude Test).

A similar set of ten correlations was calculated for the entire period 1930–1936, regardless of the year of admission. Separate correlations by years were made in addition to those for the entire period because it was felt that the latter, while valuable as a summary of what had happened, would not disclose any trends or variations that might have occurred during the years.

It will be noticed, likewise, that separate correlations were calculated between test scores and marks in each course, rather than between test scores and the average marks of all the courses. The latter procedure would have yielded a confused picture of the relationships sought. A relationship determined for the average of several courses could not be applied to any of the courses included in the average.

The number of cases that were available for each of the correlations by years and for the entire period may all be determined from Table I. The number of cases for a course in a particular year may be obtained from the column for that year, and for the entire period 1930–1936 from the last column on the right.

TABLE I.—DISTRIBUTION OF CASES BY COURSES IN EACH ACADEMIC YEAR FOR THE ENTIRE PERIOD 1930–1936

Course	1930–1931	1931–1932	1932–1933	1933–1934	1934–1935	1935–1936	Total
M-2	166	112	114	123	85	72	672
M-5	168	188	181	189	168	197	1091
M-6	120	125	111	131	108	136	731
M-7*A*	87	178	234	218	220	146	1083
M-7*B*	56	134	188	178	182	108	846
Total	597	737	828	839	763	659	4423

RESULTS AND CONCLUSIONS

The results obtained in this investigation are presented in Tables II–V, inclusive. Tables II and III contain the detailed results by years and Tables IV and V, the corresponding summary results for the entire period 1930–1936. Tables II and IV each show the coefficients of correlation between marks in each of the courses and scores on the verbal and mathematical sections; Tables III and V show the means of the course marks and the means of the scores on the verbal and mathematical sections. Tables II–V, inclusive, contain certain

abbreviations the meanings of which are necessary in order to understand the tables. The meanings of these abbreviations are given here.

M = mathematics
VS = verbal section
MS = mathematical section
r = coefficient of correlation
PE = probable error of coefficient of correlation
X' = mean of the course marks
Y' = mean of the scores on the verbal section or the mathematical section of the Scholastic Aptitude Test.

TABLE II.—CORRELATIONS BETWEEN MARKS IN THE MATHEMATICS COURSES AND SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR EACH ACADEMIC YEAR IN THE PERIOD 1930–1936

Course	1930–1931				1931–1932				1932–1933			
	VS		MS		VS		MS		VS		MS	
	r	PE	r	PE	r	PE	r	PE	r	PE	r	PE
M-2	.08	.05	.48	.04	.13	.06	.48	.05	.30	.06	.49	.05
M-5	.29	.05	.60	.03	.31	.04	.50	.04	.29	.05	.57	.03
M-6	.36	.05	.50	.05	.35	.05	.33	.05	.16	.06	.45	.05
M-7*A*	.10	.07	.46	.06	.19	.05	.48	.04	.22	.04	.46	.04
M-7*B*	.21	.08	.44	.07	.29	.05	.36	.05	.20	.05	.34	.04

Course	1933–1934				1934–1935				1935–1936			
	VS		MS		VS		MS		VS		MS	
	r	PE	r	PE	r	PE	r	PE	r	PE	r	PE
M-2	.20	.06	.50	.05	.22	.07	.57	.05	.06	.08	.56	.06
M-5	.36	.04	.67	.03	.29	.05	.59	.03	.17	.05	.60	.03
M-6	.26	.06	.41	.05	.31	.06	.51	.05	.13	.06	.47	.05
M-7*A*	.20	.04	.49	.04	.20	.04	.50	.03	.09	.06	.48	.04
M-7*B*	.21	.05	.46	.04	.22	.05	.42	.04	.06	.07	.34	.06

Analysis by Years.—(See Table II.)

1. The correlations on the mathematical section are higher than those on the verbal section in every course each year, with the exception

of the 1931–1932 group in Mathematics 6 (Analytic Geometry), when the two sections were about equal.

2. The correlations between the mathematical section and the mathematics courses are consistently good. The correlations for Mathematics 5 (College Algebra) are the best, all above 0.5 and very close to 0.6. The highest correlation in the whole series, 0.636, is that between the mathematical section and Mathematics 5 for the 1933–1934 group. In only three instances do the correlations on the

TABLE III.—MEANS OF THE COURSE MARKS AND THE MEANS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR EACH ACADEMIC YEAR IN THE PERIOD 1930–1936

Course	1930–1931			1931–1932			1932–1933		
	X'	VS Y'	MS Y'	X'	VS Y'	MS Y'	X'	VS Y'	MS Y'
M-2	4.5	44.0	36.5	4.2	47.5	31.0	5.0	43.5	33.5
M-5	5.0	47.5	51.0	5.0	46.5	50.5	5.2	45.0	49.5
M-6	5.5	48.5	55.5	5.4	46.5	57.5	5.4	45.0	58.0
M-7*A*	4.7	45.5	39.0	4.3	50.5	39.5	4.7	44.0	34.5
M-7*B*	4.8	44.5	40.5	4.6	50.0	42.5	4.8	44.5	36.0

Course	1933–1934			1934–1935			1935–1936		
	X'	VS Y'	MS Y'	X'	VS Y'	MS Y'	X'	VS Y'	MS Y'
M-2	4.8	44.5	32.5	5.2	48.5	34.5	5.0	40.0	30.5
M-5	4.9	44.0	48.0	5.3	46.5	52.0	5.1	44.5	52.5
M-6	5.0	46.5	55.0	5.3	48.0	60.5	5.2	46.5	57.5
M-7*A*	4.9	40.5	35.0	5.1	43.0	33.5	4.8	40.0	35.0
M-7*B*	5.1	40.5	37.0	5.5	43.5	35.5	5.3	39.5	38.0

mathematical section fall below 0.4, in Mathematics 7*B* (Elementary Analysis, continued) for the 1931–1932, 1932–1933, and 1935–1936 groups.

3. In every case, the correlations on the mathematical section are lower for Mathematics 7*B* than for Mathematics 7*A* (the second and first semesters, respectively, of Elementary Analysis). This is, in general, not true of the verbal section. In only two of the six correla-

tions between the verbal section and Mathematics 7*B* are they lower than the corresponding ones for Mathematics 7*A*.

Analysis over the Entire Period 1930–1936.—(See Table IV.)

1. The correlations on the mathematical section are higher than those on the verbal section for every course.

To determine the reliability of the difference between the obtained coefficients of correlation for the verbal and mathematical sections,

TABLE IV.—SUMMARY OF CORRELATIONS BETWEEN MARKS IN THE MATHEMATICS COURSES AND SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR THE ENTIRE PERIOD 1930–1936

Course	Verbal section		Mathematical section	
	r	PE	*r*	PE
M-2	.15	.03	.49	.02
M-5	.29	.02	.59	.01
M-6	.27	.02	.44	.02
M-7*A*	.17	.02	.46	.02
M-7*B*	.19	.02	.37	.02

TABLE V.—SUMMARY OF THE MEANS OF THE COURSE MARKS AND THE MEANS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR THE ENTIRE PERIOD 1930–1936

Course	*X′*	VS *Y′*	MS *Y′*
M-2	4.7	45.0	33.5
M-5	5.1	45.5	50.5
M-6	5.3	46.5	57.5
M-7*A*	4.7	43.5	35.5
M-7*B*	5.0	43.5	37.5

the ratio of the difference of the two coefficients to the probable error of their difference[1] was calculated for each course. In all the mathematics courses, the ratios so found were greater than four, indicating that there is complete certainty (one hundred chances in one hundred)

[1] $\frac{D}{\mathrm{PE}_{(\mathrm{diff.}r_1-r_2)}}$.

that the obtained difference represents a true difference greater than zero.[1]

2. The correlations on the mathematical section are all above 0.4, with the exception of Mathematics 7*B* (Elementary Analysis, continued).

3. The mathematical section is strongest in the case of Mathematics 5 (College Algebra), being close to 0.6.

4. The relationship between the verbal section and achievement in mathematics is low.

5. The means of the scores on the verbal and mathematical sections contain some interesting points. If the means of the scores on the mathematical section for the various courses in any year (Table III), or for the entire six-year period (Table V), are examined, it is seen that they do not remain constant, but show rather marked variations.

Mathematics 2 (Plane Trigonometry), Mathematics 7*A* (Elementary Analysis), and Mathematics 7*B* (Elementary Analysis, continued), the beginning courses in mathematics, are composed of students who have had very little previous mathematics, or who did not do well in their high-school mathematics, and of students who do not intend to take any additional mathematics. It would, therefore, be expected that such students would be lower in mathematical ability than those in more selected groups. For the students in these three courses, the means of the scores on the mathematical section are low. In Mathematics 5 (College Algebra) and Mathematics 6 (Analytic Geometry), composed of students of higher mathematical ability, in the main of engineering students and for the remainder of a more selected group from the other schools taking additional mathematics, the scores on the mathematical section rise, and rise sharply.

A similar analysis of the means of the scores on the verbal section shows that these scores do not exhibit corresponding variations. Instead, they remain more or less constant from course to course, thus substantiating in yet another way the evidence demonstrated by the coefficients of correlation, that the mathematical section is a more sensitive instrument than the verbal section in detecting groups of varying mathematical ability.

6. Considering that we are dealing with a single instrument and that we are on the college level, where the student personnel is more homogeneous than on the secondary-school level, the fact that the

[1] Garrett, Henry E.: *Statistics in Psychology and Education*, p. 135, Table XV.

correlations for the mathematical section are almost all above 0.4 indicates a considerable positive relation between the scores made by students on the mathematical section of the Scholastic Aptitude Test at the time of entering the University and their future academic success in specific courses in first-year mathematics. It must also be remembered that the obtained coefficients of correlation are lower than they otherwise would be because not all the candidates who apply for admission and take the examination are admitted, the poorer ones being rejected.[1] The more homogeneous the group, the greater is the restriction in the range of ability, and this restriction tends to lower the coefficients of correlation.[2]

7. Beginning with the June, 1936, Scholastic Aptitude Test, taken in the main by candidates for admission in September, 1936, the College Entrance Examination Board dropped the mathematical section. This move by the Board did not affect any of the students considered in this study, since the study concerned itself with students admitted between September, 1930, and February, 1936, during which time the verbal and mathematical sections did not change and were taken by all applicants.

In place of the mathematical section, the Board is offering a new series of mathematical examinations, called Alpha, Beta, and Gamma.[3] Since the Board cannot compel the acceptance of its offerings, it remains for the individual universities to decide upon their acceptance or rejection.

The results obtained on the mathematical section in this investigation show sufficient promise to indicate the desirability for the continued use of such a section in the Scholastic Aptitude Test for the guidance and teaching of students in courses in mathematics. Since the original mathematical section is no longer available, and since the

---

[1] *Twenty-Sixth Annual Report of the College Entrance Examination Board*, 1926, p. 8.

[2] Byrns, Ruth K.: "Scholastic Aptitude and Freshman Achievement." *School and Society*, Vol. XXXV, May 21, 1932, p. 718.

Crawford, A. B.: "Forecasting Freshman Achievement." *School and Society*, Vol. XXXI, January 25, 1930, p. 129.

Freeman, Frank S.: "Predicting Academic Survival." *Journal of Educational Research*, Vol. XXIII, February, 1931, p. 116.

Segel, David: *Prediction of Success in College.* U. S. Office of Education Bulletin, 1934, No. 15, p. 19.

[3] Report of the Commission on Examinations in Mathematics, *Mathematics Teacher*, Vol. XXVIII, March, 1935, pp. 154–166.

College Entrance Examination Board is now offering the new mathematical examinations Alpha, Beta, and Gamma in its place, it is suggested that these new examinations be given a trial. It would remain, however, for a future study similar to this one, by comparing the results of such a study with those found here, to determine whether the new mathematical examinations are more closely related to marks in mathematics than the original mathematical section.

ROSANOFF, HANDY, AND PLESSET ON THE ETIOLOGY OF MENTAL DEFICIENCY; A CRITICAL APPRAISAL

FRANK K. SHUTTLEWORTH

Institute of Human Relations, Yale University

The recent monograph by Rosanoff, Handy, and Plesset on the etiology of mental deficiency[1] is distinguished for the conceptual framework of its attack, for the care involved in the collection of a large unselected series of cases, for the richness and wealth of new quantitative data and case records which are reported *in extenso*, and for the ingenuity employed in testing a wide range of important hypotheses. It is also distinguished for its failure to subject its findings and conclusions to any critical appraisal. This paper attempts to supply the necessary critical appraisal. The three major conclusions will be considered seriatim.

HEREDITARY FACTORS IN MENTAL DEFICIENCY

Of three hundred sixty-six twin pairs with mental deficiency (defined as IQ under 80) in one or both of each pair, one hundred twenty-six are classified as monozygotic, one hundred one as same sex dizygotic, and one hundred thirty-nine as opposite sex dizygotic. Omitting cases for which IQ's are not available or can not be estimated, the authors report the following findings: among one hundred eleven monozygotic pairs, among eighty-seven same sex dizygotic pairs, and among one hundred sixteen opposite sex dizygotic pairs the IQ's do not differ by more than five points in 64.9 per cent, 20.7 per cent, and 14.6 per cent of the pairs respectively. The authors contrast the percentage of close agreement among the monozygotic pairs (64.9 per cent) with the percentage of close agreement among the two dizygotic groups combined (17.2 per cent), the difference being 47.7 per cent. From this contrast they conclude (page 117) "that scarcely more than one-half of the cases are of hereditary origin." Many serious difficulties are involved in this conclusion.

First, it may be noted that the authors present distributions of IQ differences for a total of three hundred fourteen twin pairs, whereas the writer can find in the case histories only two hundred forty-seven

[1] Rosanoff, Aaron J., Handy, Leva M., and Plesset, Isabel R.: "The Etiology of Mental Deficiency With Special Reference to Its Occurrence in Twins." Psychological Monographs, Vol. XLVII, No. 4, 1937, pp. 137.

pairs both of which have recorded IQ's. The writer does not question the validity of the estimated IQ's which are evidently involved, but it is unfortunate that these estimated IQ's were not quantified, labelled as such, and systematically reported in the case histories.

Second, there is the difficulty of determining what proportion of monozygotic twins truly resemble each other when allowance is made for the unreliability of the Stanford-Binet. From data of Lincoln[1] the writer estimates the probable error of the difference between two individual IQ's in the range below 80 at 5.1 IQ points. This is a very severe standard, equivalent to saying that the probable error of an individual IQ is only 3.6 points. Further, no allowance has been made for the fact that 51.5 per cent of the individual twins present one or more complications such as paralysis, epilepsy, psychosis, behavior problems, delinquency, or criminality. These complications tend to increase the unreliability of obtained IQ's. The percentage distribution of IQ differences to be expected from this estimated unreliability of the test together with the obtained percentage distribution of differences in the monozygotic pairs taken from Table III are as follows:

	Percentage distribution of IQ differences					
	0–5	6–10	11–15	16–20	21–25	Over 25
Due to unreliability.....	53.3	30.2	12.4	3.4	0.6	0.1
Monozygotic pairs.....	64.9	18.0	6.3	2.7	3.6	4.5

Obtained differences between monozygotic pairs exceed expectations due to unreliable IQ's by three per cent in the interval 21–25 and by 4.4 per cent in the interval over 25, or a total excess of 7.4 per cent. Or, allowing for the unreliability of the Stanford-Binet, the monozygotic twins are so very similar that with perfect tests 92.6 per cent might well prove to have identical IQ's.[2]

---

[1] Lincoln, Edward A.: "Stanford-Binet IQ Changes in the Harvard Growth Study." *Jour. Appl. Psychol.*, Vol. XX, 1936, pp. 236–242.

[2] Although the distributions are not such as to justify the calculation of correlations, the close resemblance of the monozygotic twins may, nevertheless, be verified by this method, the correlation between their IQ's being .896. This is as high as the correlations obtained between retest IQ's with unselected children.

Third, there is the very great difficulty of constructing a comparable measure of true resemblance for the dizygotic pairs. The same procedure, as outlined above, can not be applied to two such diverse groups as monozygotic and dizygotic twins when propositi are selected from the extreme tail of a normal distribution. This point may be illustrated by assuming the availability of a large unselected population of monozygotic and of dizygotic twins and the selection from each population of two groups of propositi, one with IQ's between 95 and 105 and the other with IQ's under 80. Now, the IQ differences between monozygotic pairs for propositi having IQ's between 95 and 105 will be very similar to the IQ differences between monozygotic pairs for propositi having IQ's below 80, the second set of differences being only slightly larger. But, when the same procedure is applied to dizygotic twins the two sets of IQ differences will be radically dissimilar, the second set of differences being very much larger.

Fourth, let the proportion of pairs who truly resemble each other be indicated by the letters a, b, and c, respectively, for monozygotic, dizygotic, and sibling pairs. Then, it should be obvious that the difference between monozygotic and dizygotic pairs, $a - b$, can not give even an approximate estimate of the proportion of mental defective cases to be attributed to hereditary factors because this ignores the fact that the resemblance of dizygotic twins is to be attributed in part to hereditary factors. The proper formula is $a - b + c$; namely, the proportion of true resemblance among sibling pairs must be added to the difference between monozygotic and dizygotic pairs.

In summary of the foregoing remarks, the methods of measuring resemblance employed by the authors are valid only for the purpose of demonstrating that monozygotic twins resemble each other more closely than do dizygotic twins. Their methods, however, are wholly invalid to support their conclusion that "scarcely more than one half" of the cases of mental deficiency are of hereditary origin. Indeed, given the available data, the writer is quite unable to see how any valid conclusion can be drawn concerning the proportion of mental defective cases to be attributed to hereditary factors.

CEREBRAL BIRTH TRAUMA

The second major conclusion is that cerebral birth trauma is "a factor of the highest importance in the etiology of mental deficiency" (page 131). In the light of the first conclusion it is clearly implied,

though not explicitly stated, that cerebral birth trauma is a more important factor than heredity. Since both the evidence and the argument are rather involved, the writer has numbered and catalogued under eight major headings the facts cited by the authors in support of this conclusion. It should be noted that their argument is actually focused on mental deficiency in twins rather than in singly-born subjects although this important distinction is only incidentally mentioned (page 117).

1. The existence of relative mental deficiency among monozygotic twins of superior intelligence "is demonstrated not infrequently" (page 130). Of thirty-four pairs of IQ's from monozygotic twins of superior intelligence, four pairs show differences of 13, 14, 15, and 32 IQ points. "Surely" these four differences "must be attributed to prenatal or intranatal factors, and not to hereditary or germinal factors" (page 109). Relying once more on Lincoln's data, the writer finds all four differences to be within the range of what might be expected solely from the unreliability of the Stanford-Binet.

2. Prematurity and underweight condition at birth are associated with feeble-mindedness. (*a*) Premature birth or underweight condition, or both, was found in 21.3 per cent of a group of one hundred twenty-two institutionalized feeble-minded and in only 3.9 per cent of a series of nine thousand seven hundred eighty-two normal control birth records (pages 124–125). (*b*) In a series of one hundred forty-six cases either prematurely born or underweight at birth, or both, 10.3 per cent had IQ's under 76, whereas among Terman's nine hundred five unselected cases this proportion is only 2.6 per cent (pages 126–127). The writer finds this difference to be statistically reliable (more than four times its probable error) but the difference in mean IQ as calculated by the writer is only 0.2 points with a PE of 1.1 in favor of the unselected cases.

3. Twin births are subject to greater hazards than single births. (*a*) Premature births or underweight condition at birth or both are more common among twins (page 119). This is certainly true as to underweight condition, but no evidence is given as to prematurity. (*b*) Citing two studies, it is stated that mental deficiency is more common among twins (page 119). On examining these studies the writer finds no evidence for this statement unless it be a reference to a study by Humm who found that among cases selected for mental deficiency there is greater concordance of findings for dizygotic twins than for singly-born siblings.

4. Monozygotic twins are subject to greater hazards than dizygotic twins (same and opposite sex combined). (*a*) In the year 1930 of monozygotic twin births (total individuals) the authors estimate that eleven and seven tenths per cent were stillborn whereas only six per cent of opposite sex twins were stillborn (page 119). (*b*) Severe mental deficiency as indicated by IQ's under 50 occurs in 20.2 per cent of monozygotic individuals and in only 10.4 per cent of dizygotic individuals (page 120). (*c*) "It is highly probable that among surviving twins the incidence of mental deficiency, infantile palsies, epilepsy, and allied conditions attributable to cerebral birth trauma, is higher in the monozygotic, as compared with the dizygotic group" (page 120). No evidence to support this speculation is cited save that already recorded under (*b*). From Table IV and Tables IX plus XIV the writer finds the incidence of epilepsy to be 8.3 per cent and 7.7 per cent, of infantile palsies four per cent and 1.5 per cent and of psychotic disease 4.4 per cent and 3.1 per cent in each case, respectively, among monozygotic and dizygotic twins. None of these differences is statistically reliable. The relevance of (4*a*), (4*b*) and (4*c*) to the problem is not clear to the writer, nor do the authors offer any explanation of these unusual phenomena. In explanation the writer ventures to suggest that monozygotic twins have only half as much original germ plasm to begin with. If this is the correct explanation of (4*a*) and (4*b*), it follows that monozygocity *per se* is a unique etiological factor in mental deficiency, and *ipso facto* a wholly irrelevant factor in the etiology of mental deficiency among singly-born cases.

5. The incidence of complications increases with the severity of the mental deficiency. (*a*) Of one hundred forty individual twins with IQ's of 85 or higher only 12.8 per cent present complications such as infantile palsies, epilepsy, psychotic manifestations, child behavior difficulties, juvenile delinquency, and adult criminality whereas of five hundred ninety-two individual twins with IQ's of 84 or lower 43.7 per cent present these complications (page 120). (*b*) Of individual twins with IQ's of 50 or higher only 0.8 per cent were observed with infantile palsies, only 4.1 per cent with epilepsy, and only 2.5 per cent with psychotic disease, whereas of individual twins with IQ's of 49 or lower the comparable proportions were 17.9 per cent, 33.7 per cent, and 9.5 per cent, respectively (page 121). The number of cases is not stated, but they are probably six hundred thirty-one and one hundred one. Hence, save for psychotic disease, the differences are statistically reliable.

6. The importance of non-hereditary factors increases with the severity of the mental deficiency and with the presence of complications. (*a*) "Certainly our material suggests—that non-hereditary mental deficiency comprises the bulk of the cases on the level of idiocy and imbecility" (page 120). This statement follows directly after that quoted under (4*c*), as if implying that cerebral birth trauma is responsible for the bulk of the cases on the level of idiocy and imbecility. No evidence is cited directly in support of the quoted statement, although that under (5*b*) may be pertinent. However, evidence under (4*b*) and the writer's comments thereon suggest that the non-hereditary factor of monozygocity *per se* may be an etiological factor at the level of idiocy and imbecility, but, if so, this would contribute nothing to the understanding of mental deficiency in general. (*b*) When monozygotic pairs are divided into those with IQ's over 50 and those with one or both IQ's under 50 the proportions of pairs showing close resemblance in IQ are 68.8 per cent and 54.5 per cent, respectively (page 117). Perhaps, it is these data to which the quotation under (6*a*) refers, but the difference is not statistically reliable. The difference may be due to selection of *both* having IQ's over 50 for comparison with *one or both* having IQ's under 50. Comparable data for dizygotic pairs are not reported. (*c*) When twin pairs are divided into those without any complications and those with complications in one or both of each pair the proportions showing close resemblance in IQ are 72.6 per cent and 47.6 per cent for monozygotic pairs, 28.3 per cent and 11.9 per cent for same sex dizygotic pairs, and 22.1 per cent and 3.5 per cent for opposite sex dizygotic pairs (pages 39, 63, 95, 117). The number of cases is not stated but the writer estimates that these contrasts are reliable only for the monozygotic and opposite sex dizygotic pairs. The writer strongly suspects that these contrasts may be due to the greater unreliability of the Stanford-Binet when applied to cases with the indicated complications.

7. The concordance of findings for infantile palsies, epilepsy, and psychotic condition is greater among monozygotic than dizygotic twin pairs. "This suggests that hereditary factors play a part even in cases such as those of infantile cerebral palsies, in which the essential etiology definitely consists almost exclusively in cerebral birth trauma" (page 123). Since the authors merely refer to the basic tables, the writer summarizes the facts as follows: (*a*) Of one hundred twenty-six monozygotic pairs both members have infantile palsies in six cases, both have epilepsy in sixteen cases, and both have psychotic disease

in six cases; a total of twenty-eight cases or 22.3 per cent, whereas the comparable figure for all dizygotic pairs is only 5 per cent. The difference is statistically reliable. The difference for epilepsy considered separately is also reliable. (*b*) Of forty-two propositi among monozygotic twins having one of these three complications, the same condition is found among 66.7 per cent of the second members of the pairs whereas the comparable figure for eighty-four propositi among dizygotic twins is only 14.3 per cent. This difference is reliable. Comparable proportions of concordance for epilepsy among twenty-one monozygotic propositi and forty-nine dizygotic propositi are 76.1 per cent and 12.1 per cent, respectively. This is also reliable. Of ten monozygotic and twelve dizygotic propositi with infantile palsies the concordances are sixty per cent and zero. These are extreme differences proving, at least to the writer, that hereditary factors are involved. The writer suggests that these facts together with those under (5*b*) would be interpreted by many geneticists as evidence of some form of underlying gene-linkage.

8. The concordance of findings for mental deficiency and for epilepsy is greater in dizygotic than in singly-born sibling pairs. (*a*) These concordances in the case of mental deficiency are 53.3 per cent and 16.5 per cent, respectively (page 118). (*b*) In the case of epilepsy the concordances are, respectively, 10.7 per cent and from 1.1 per cent to 4.1 per cent (page 119). The number of cases involved in these two statements is not given and the writer has been unable to verify all of the foregoing figures, but from his own analyses of original data the differences are clearly reliable.

9. Finally, the writer would add that the incidence of epilepsy or infantile palsies, or both, among mental defectives as calculated from summary tables is 15.5 per cent.

The result of this examination of seventeen specific issues of fact discredits the significance of eight (1, 2*b*, 3*a*, 3*b*, 4*c*, 6*a*, 6*b*, and 6*c*) and marks two as irrelevant (4*a* and 4*b*). We are left with seven verified items (2*a*, 5*a*, 5*b*, 7*a*, 7*b*, 8*a* and 8*b*). Of these the proof for two (7*a* and 7*b*) was supplied by the writer and two (2*a* and 5*a*) are not directly relevant. The argument for cerebral birth trauma boils down to two facts. First, the incidence of epilepsy and infantile palsies, which are almost always associated with cerebral birth trauma, increases markedly with the severity of the mental deficiency (5*b*). Hence, cerebral birth trauma would seem to be an etiological factor in the more severe grades of mental deficiency. Second, dizygotic

twin pairs resemble each other in the occurrence of epilepsy and of mental deficiency to a greater degree than do ordinary sibling pairs (8*a* and 8*b*). This proves that non-hereditary factors are important in the etiology of these conditions and, hence, suggests that the non-hereditary associated factor of cerebral birth trauma is also important. To balance this evidence two additional facts are equally clear. The incidence of epilepsy and of infantile palsies, and of both, among mental defective twins is only 15.5 per cent (9). Hence, on the above argument, cerebral birth trauma is a probably important factor in only 15.5 per cent of the cases. Second, monozygotic twin pairs resemble each other in the occurrence of epilepsy, of infantile palsies, and of mental deficiency to a much greater degree than do dizygotic twin pairs (7*a* and 7*b*). This proves that hereditary factors are important in the etiology of these conditions and, hence, suggests that hereditary factors are also important in the etiology of cerebral birth trauma itself.

The writer would restate the findings as follows: Cerebral birth trauma is a probably important factor in fifty per cent of the lower grades of mental defective twins with IQ's under 50, a probably important factor in five per cent of the higher grades of mental defective twins, and a probably important factor in less than five per cent of the higher grades of singly-born mental defectives. That is, only a small fraction of mental deficiency in general can be attributed to cerebral birth trauma.

SEX FACTORS IN INTELLIGENCE

The authors present three bodies of data each of which proves to them that sex is a factor in intelligence, the girls being slightly favored. First, they summarize the evidence from one hundred thirty-nine pairs of opposite sex twins selected for mental deficiency in one or both pairs. IQ's under 85 are reported for one hundred eleven males and eighty-seven females: "A relative excess of males amounting to 27.6 per cent." They add, "The conclusion is hardly to be evaded, that sex is an etiological factor in mental deficiency" (page 96). The writer concludes, instead, that the authors found more males than females in the institutions which they contacted and, hence, that these data have not the slightest bearing on the problem.

Second, the authors present data on two hundred thirty-three pairs of opposite-sex twins of normal intelligence selected from public-school populations in southern California. Excluding six cases with identical

IQ's, girls surpass their twin brothers in 55.1 per cent and fall below in 44.9 per cent of the comparisons. The obtained proportions depart from chance by 5.1 per cent. The writer finds the probable error of 5.1 per cent to be 2.2 per cent; hence, the difference is not reliable. Since these data are reported in extenso, the writer has calculated the means, standard deviations, and the correlation between the pairs of IQ's. The IQ's of the girls average $1.52 \pm .65$ points higher; hence, again, the difference is not reliable. It should be noted here that the authors' method of stating their results involves two illicit exaggerations. In the first place, calculation of the difference between 55.1 per cent and 44.9 per cent is improper because these percentages are not independent events. The authors go still further, calculating the relative excess of higher IQ's in the girl (10.2 per cent divided by 44.9 per cent) and by this process inflate an actual difference of only 5.1 per cent into a relative excess of 22.5 per cent.

The third body of data consists of IQ's on seven hundred thirteen pairs of opposite-sex, singly-born siblings of normal intelligence selected from eight hundred eighty sibling groups. Excluding twenty-one pairs with identical IQ's, girls are superior in 54.5 per cent of the comparisons. This departs from chance by 4.5 per cent, but the writer finds its probable error to be 1.3 per cent; hence, again, the difference falls short of reliability. The authors also report the average IQ of all one thousand ten boys and of all one thousand seven girls in these eight hundred eighty sibling groups to be 105.6 and 107.9, respectively. The writer finds the probable errors of these averages to be .306 and .298. By the conventional formula the probable error of the difference is .43; hence, the obtained difference of 2.3 IQ points is more than five times its own probable error. There are at least four difficulties even with these results. First, the authors should report average IQ's and standard deviations for pairs of opposite-sex siblings rather than for all boys and all girls. Second, they should report the number of different individuals of each sex involved in their seven hundred thirteen pairs. Since each boy was compared with each of his sisters, it is obvious that some individuals are involved in several pairs. Third, the correlation between the IQ's of these pairs is essential to statistical analysis. Fourth, it would be desirable to report findings separately for first-born boys *versus* later born girls and for first-born girls *versus* later born boys. However, combining the data from two hundred thirty-three opposite-sex twin pairs and seven hundred thirteen opposite-sex sibling pairs, and excluding twenty-seven pairs

with identical IQ's, the writer finds that girls are superior in 54.6 per cent of the nine hundred nineteen comparisons. The departure from chance of 4.6 per cent has a probable error of 1.1 per cent; hence, the difference is significant in a technical and statistical sense.

The writer has checked the trend of these findings with IQ's for opposite-sex twins published *in extenso* by Merriman[1] and by Wingfield.[2] Although the number of cases is very small they support the findings of Rosanoff, Handy, and Plesset for younger children. Over ages ten and twelve, however, the data of Merriman and Wingfield suggest that boys may be equal or superior. This suggests in turn that the findings of Rosanoff, Handy, and Plesset may represent a developmental phenomenon rather than a true sex difference. To test this point the authors have courteously supplied the writer with the chronological ages of their opposite-sex twin pairs. Selecting two hundred fourteen pairs who were tested within one year of each other and plotting the IQ difference against age shows quite clearly that girls maintain their superiority equally well throughout the age range from three to seventeen. The difference in favor of girls, accordingly, can not be attributed to developmental factors but must be a true sex difference.

Finally, it needs to be pointed out that sex is an extremely small factor determining individual differences in intelligence. Considering the two hundred thirty-three twins, for example, the standard deviation for all four hundred sixty-six cases is 14.94 with sex operating, whereas when sex differences in average IQ are eliminated the standard deviation is reduced only to 14.90. Squaring these figures it appears that sex factors account for only one-half of one per cent of the IQ variance or individual differences. Similarly, sibling data show that sex accounts for only 1.3 per cent of the IQ variance. These results are in harmony with the observation that a distinction must be made between technical statistical significance and significance in any broader sense.

---

[1] Merriman, Curtis: "The Intellectual Resemblance of Twins." *Psychological Monographs*, Vol. XXXIII, No. 5, 1924, p. 58.

[2] Wingfield, Alex H.: *Twins and Orphans: The Inheritance of Intelligence.* London and Toronto: Dent & Sons, 1928, p. 127.

A STATISTIC FOR CIRCULAR SCALES

ROBERT T. ROSS

Stanford University

The statistic of linear scales has received much adequate attention in the past and its rules and procedures have been defined with considerable precision. There remain, however, a small but significant number of problems which are not amenable to treatment by linear methods, but which demand clarification; we may group these problems under the general classification of circular scales.

The problem may perhaps be most clearly defined by an illustration. Suppose a circular pattern of eight lights to be so arranged that one light is at the top (U) of the circumference of a circle, one at the bottom (D), one at the right (R), one at the left (L), and four equally spaced between these four (UR, UL, DR, DL). The pattern is then of eight lights equally spaced around the circumference of a circle so that four pairs are on opposite ends of four diameters as shown in Fig. 1.

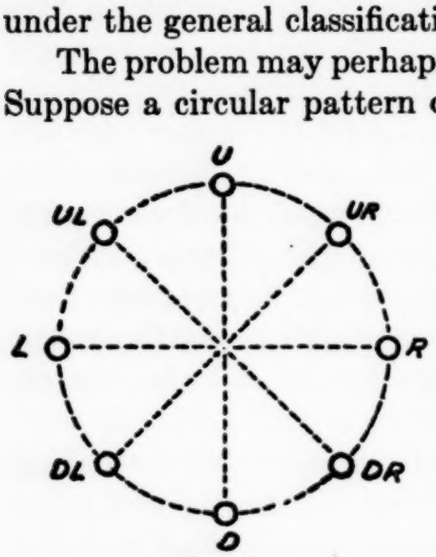

FIG. 1.

Now suppose that the bulbs are flashed on and off, one at a time, in some prearranged order, and that at the end of ten flashes an audience of eighty people is asked to guess what bulb will flash next. We wish to determine the central tendency of these responses.

The extremes of the possibilities of responses would be that all eighty subjects would guess the same light, say (U), or, that there would be ten guesses for each light. In the first case we could say that if the experiment were repeated, the most probable guess would be for U; in the second case we could not state which way the vote would go for any individual, although for the group we would suppose that it would be equally distributed.

These hypothetical cases are, of course, extreme, and in any real situation we would expect the responses to group about some central point and to spread outward from there with decreasing frequency. The difficulty encountered is that the "ends" of the distribution may overlap on the circular scale. When this happens, difficulties arise which may be solved in various ways, some of which will be discussed.

A. TRANSFORMATION TO A LINEAR SCALE

The essential feature of transforming a circular scale to a linear one is the assignment of linear scale values to the various locations on the circular scale. In the following discussion we shall make the usual assumptions of continuity of scale in referring back to an arbitrary designation of *U* as zero, *UR* as one, *R* as two, *DR* as three, *D* as four, *DL* as five, *L* as six, and *UL* as seven.

Using this arbitrary system, suppose thirty guesses fall at *UL*, thirty at *U*, and twenty at *UR*. Using the linear scale values and averaging, we obtain an average value of approximately 2.8, which is obviously absurd. We may avoid this absurdity by moving the arbitrary origin to *UL*, whence an average gives us a value on the original scale of 7.9, which is reasonable.

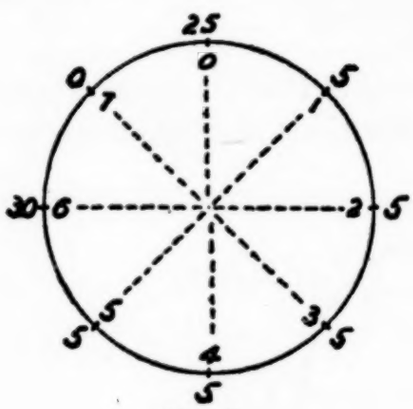

FIG. 2.

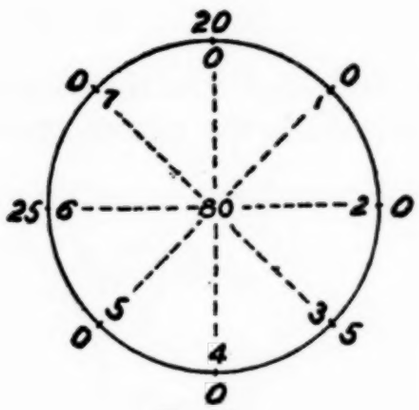

FIG. 3.

An extension of this argument might lead us to assume that we can start an arbitrary linear scale wherever there is a break in the circular scale. An examination of Fig. 2, however, indicates that this general rule may lead to absurdity, for a use of linear scale values leads to an average of 3.2, which is untenable.

One is tempted, in this extremity, to use what might be called a "cancellation" method and argue that, since the five judgments at *UR*(1) are exactly counterbalanced by the five at *DL*(5), we might consider that these values cancel each other, and proceed in a like manner to cancel all values at opposite ends of the same diameters. We arrive by this method at the distribution shown in Fig. 3. The "eighty" in the center of the circle is the original *N*. But even this arrangement is unsatisfactory, since the position of the arbitrary origin changes the value of the central tendency.

The only rule for transformation seems to be the logical one that any circular scale may be transformed into a linear one *provided the frequencies are all to be found on one side of the extremes of a given diameter.* This follows, somewhat tenuously to be sure, from the consideration that the opposite ends of a linear scale are "opposite" in much the same sense as are the ends of a diameter. But a linear scale can accommodate only two opposite ends, whereas a circular scale can accommodate an infinite number. In the special case, however, where the circular scale accommodates only one pair of opposites (two ends), it may be successfully transformed to a linear scale. If treatment of a circular scale by the cancellation method results in a distribution which complies with the requirements of transformation, presumably the resultant scale may be transformed.

B. VECTORIAL METHOD

It is obvious from the foregoing discussion that there are a number of instances in the various possible distributions of frequencies on the

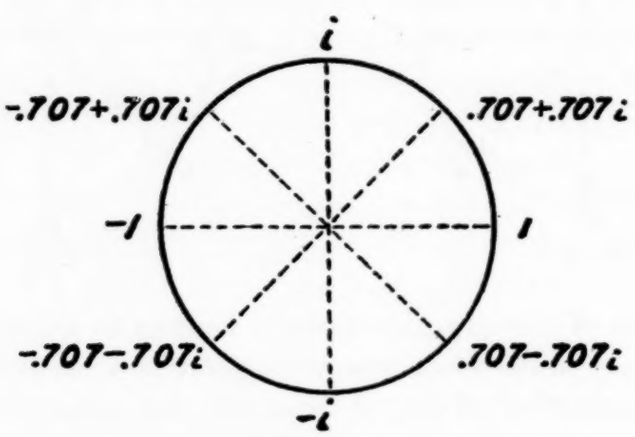

FIG. 4.

circular scale which cannot be treated by linear transformation methods. The primary difficulty seems to center around the fact that linear scales have only two opposite ends. A solution of the problem of the circular scale demands, therefore, a statistic without this limitation. Such a solution may be found in the vectorial method.

In using vectors, we assume that the values on the circular scale are on the circumference of a circle of unit radius and represent these vectors in the notation of complex numbers $(a + bi)$. For con-

venience we may express the vectors in polar coördinates whenever this seems desirable.[1]

According to this notation, the points on the scale shown in Fig. 1 correspond to the unit vectors shown in Fig. 4. In terms of polar coördinates, $r = 1$ for all cases and θ varies from zero degree to three hundred sixty degrees by steps of forty-five degrees beginning with R and proceeding counterclock wise.

We may now turn our attention to the problem presented in Fig. 3 and attempt its solution by the vectorial method. We first assume that each unit of frequency is equivalent to scaler multiplication of the appropriate vector by unity. The three vectors represented in Fig. 3 may then be appropriately multiplied and added, whence we obtain

$$\begin{array}{rr} 20i & (V) \\ 3.535 - 3.535i & (DR) \\ -25 & (L) \\ \hline -21.465 + 16.465i & \end{array}$$

This resultant corresponds in polar coördinates to a vector of length

$$\begin{aligned} r &= \sqrt{(21.465)^2 + (16.465)^2} \\ &= 27.05 \end{aligned}$$

and angle

$$\begin{aligned} \theta &= \tan^{-1}\frac{16.465}{-21.465} \\ &= 142.5^\circ \end{aligned}$$

A solution of the original problem shown in Fig. 3 gives the same result.

The vectorial method provides us, then, with an angular value (θ) indicative of the position of the central tendency, and with a scalar value (r), which varies from a maximum of N when all of the cases fall at the same point on the scale, to zero when they are equally distributed. If we choose an index P such that

$$P = \frac{r}{N}$$

we then have a value which varies from zero to unity regardless of the

[1] Reference to any standard college algebra will provide the reader with a discussion of the fundamentals of this notation.

value of N and which gives us a measure of the precision of the central tendency, unity indicating complete consistency and zero indicating indeterminacy. Some conception of the meaning of P may be had by considering its various values for *equal frequencies* separated by various angular differences as shown in Table I.

TABLE I

ANGULAR SEPARATION	P
0°	1.0
45°	0.9
90°	0.7
135°	0.4
180°	0.0

It would seem that the vector method provides a satisfactory solution for the central tendency and at the same time provides a measure of the precision of this value.

THE CONSTRUCTION OF CIRCULAR SCALES

Although the circular scale is not common, its very rarity may lead to errors in construction. The two major criteria which a circular scale must fulfill are: (1) It must be continuous in fact as well as arrangement, and (2) the values opposed on diameters must be true opposites. Here, as elsewhere in statistics, the use of the statistical method cannot compensate for errors in scale construction. The statistical method is absolutely impartial as to what values are put into it, the meaning of such measures as it delivers depends entirely on the logic behind the scale and the method.

An objective test for either continuity or opposition is not easy. With regard to continuity, the test of constant differences may be applied, by which we mean that between two points on the scale separated by a given interval, a constant difference must be present regardless of the position of the points. But this criterion is difficult of application. Obviously, the example cited in Fig. 1 is a circular scale problem, and certainly there is more difference between seven and zero than between any other two equally spaced points; but disregarding the numerical scale, to find a test to show that the same difference exists between *UL* and *U* as between *D* and *DL*, other than our own psychological judgment, is well-nigh impossible. It is felt, however, that a careful evaluation of a scale will show whether or not it may be considered circular.

The second criterion is likewise difficult to measure. Consider, for example, the familiar color circle, an array of colors passing from red through orange, yellow, green, blue, violet, and purple, and back to red again. This scale is sufficiently continuous to satisfy the criterion of continuity, but before a vectorial method can be used with it, it must be demonstrated that the scale has been so arranged that a color diametrically opposed to another may be thought of as logically its opposite. This demands, of course, a definition of what logically

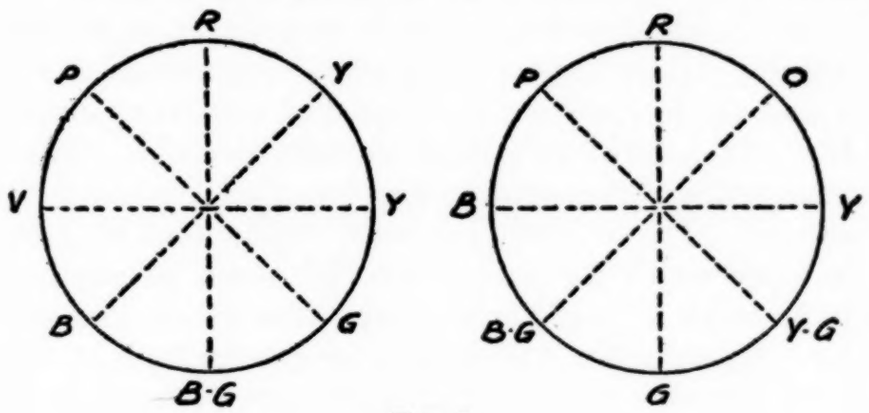

FIG. 5.

opposite colors are. Two possible scales are shown in Fig. 5. It is probable that if each of the opposite colors in the right-hand scale were defined as complementary, the scale would be acceptable to most workers. The left hand scale, while continuous, lacks satisfactory "opposition."

SUMMARY

The problem of the circular scale is defined and the difficulties of transformation to linear scales are discussed. A vectorial method of solution is presented which gives a measure of central tendency and a measure of precision in terms of the vectorial angle and scalar magnitude. The criteria of "continuity" and "opposition" in circular scale construction are pointed out and discussed.

INTELLIGENCE TESTING BY DRAWINGS

JOSEPH MILLER

Public Schools, Wilkes-Barre, Pa.

Binet[1] used in his scale two drawings, a square and a diamond. The square was placed in the fourth year and the diamond in the seventh year. In discussing these tests, Terman[2] suggests that: "It would be worth while to devise and standardize still more complicated figures." The following study is an attempt to do this. The purposes were: (1) To standardize a series of drawings of increasing difficulty and (2) to ascertain the degree of correlation between the ability to copy the drawings and general intelligence as measure.

The final series of twenty drawings (see Fig. 1) was selected from forty drawings. The selection was made on the basis of trying them out. Drawings which the author thought would be relatively easy proved to be much more difficult to copy than others that seemed to be simpler. Drawing No. 8 seemed to be more difficult than No. 15; drawing No. 19 easier to copy than many which precede in the final arrangement. The final sequence of the drawings as they appear in Fig. 1 was arranged in order of difficulty according to a percentile scale of the results of three hundred pupils of various grades.

Binet had similar difficulties in placing his drawings at the proper age. In his 1911 revision he shifted the copying of the diamond to the sixth-year level. In the final edition, however, this test is given to a child of seven because only half of the six-year-olds succeeded in making a creditable copy.

The final series of twenty drawings (Fig. 1) was presented to one hundred pupils in each grade (age constant for a given grade) from nursery school to the seventh grade. A total of nine hundred pupils was tested. The figures were drawn in heavy black lines on white cardboard (15 × 11). The pupils were given pencils and plain white paper with the instruction: "I want you to make one just like it." During the drawing the children were admonished again and again to be careful and to draw exactly as they saw. They were not restricted in their time; the samples were held before them until they finished their drawings. As soon as one drawing was completed the next was presented.

---

[1] *The Stanford Revision of the Binet-Simon Tests* (1916 and 1922 ed.).

[2] Terman, Lewis M.: *The Measurement of Intelligence*, 1916, p. 157.

Many children have an aversion to drawing on request and have to be urged. Pupils claim that they do not known how to draw and that they never liked to draw. This is especially true of children in the upper grades. Terman[1] made similar observations.

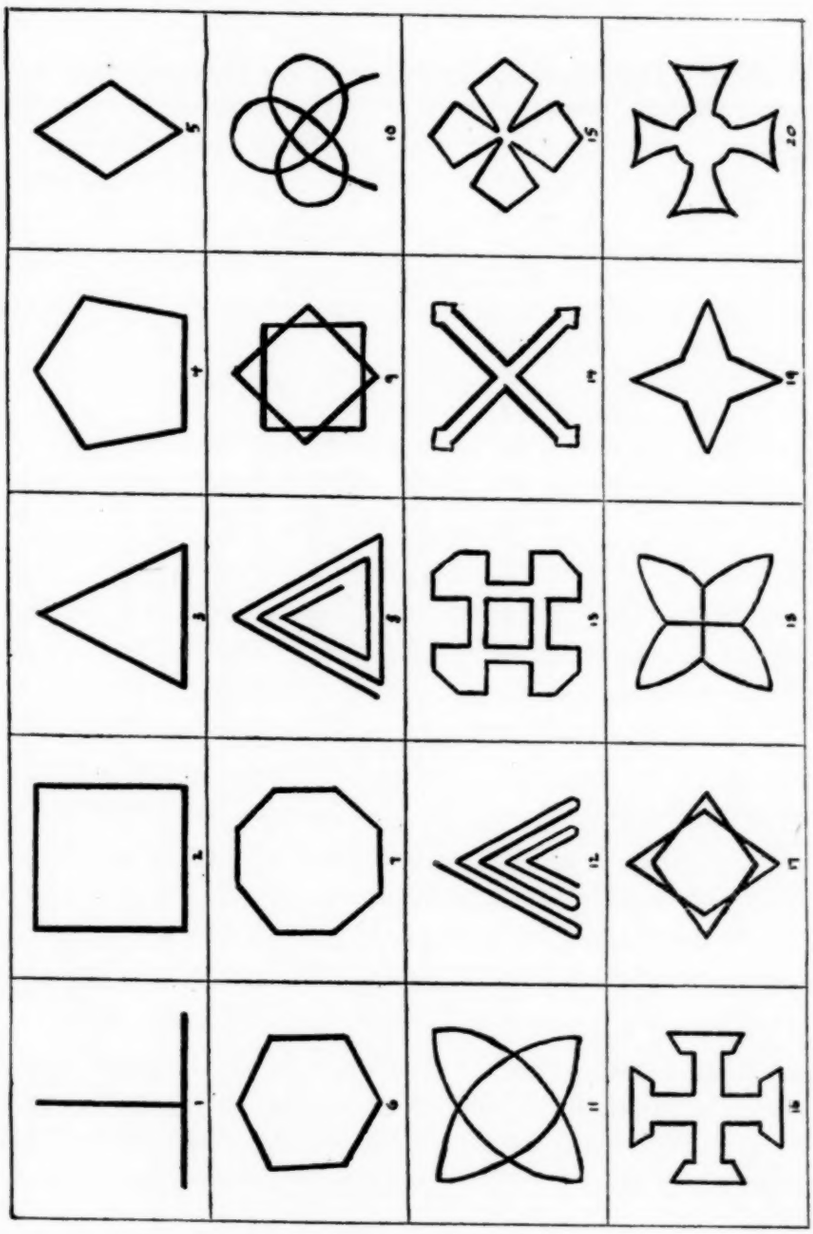

FIG. 1.—Final series of drawings.

As in the Binet-Simon intelligence scale, the scoring in this test was very simple and liberal. Credit of one was given when the drawing was correct, *i.e.*, if there was a proper relationship between the lines and if the right angles were fairly well preserved. For credit it was

[1] Terman. *Op. cit.*, p. 155.

not required that the lines be completely straight and corners entirely closed. Size was wholly disregarded in the scoring.

Drawings from nine hundred children, ranging from four to twelve years, were used in the standardization of the scale. The subjects were selected from various schools in the public schools of Wilkes-

TABLE I.—DISTRIBUTION OF NUMBER OF DRAWINGS BY AGE

	Age								
	4	5	6	7	8	9	10	11	12
	Number of drawings								
0	29	3	2						
1	66	27	2						
2	5	24	17	3					
3	.....	25	38	18	11	1	2		
4	.....	16	21	13	9	5	2	1	
5	.....	4	11	13	9	8	7	4	2
6	.....	2	5	20	12	9	6	3	3
7	.....	.....	1	11	17	7	9	5	4
8	.....	.....	3	7	9	10	8	7	6
9	.....	.....	.....	9	16	15	12	6	4
10	.....	.....	.....	3	9	12	10	9	9
11	.....	.....	.....	1	8	11	11	10	11
12	.....	.....	.....	2	.....	9	7	16	12
13	.....	.....	.....	.....	.....	6	8	14	11
14	.....	.....	.....	.....	.....	6	4	11	5
16	.....	.....	.....	.....	.....	1	9	5	9
16	.....	.....	.....	.....	.....	.....	4	3	6
17	.....	.....	.....	.....	.....	.....	1	2	6
18	.....	.....	.....	.....	.....	.....	.....	4	6
19	.....	.....	.....	.....	.....	.....	.....	.....	4
20	.....	.....	.....	.....	.....	.....	.....	.....	2
Total cases....	100	100	100	100	100	100	100	100	100
Mean.....	.76	2.46	3.30	5.56	6.99	9.07	10.01	11.40	12.65
SD.....	.50	5.08	4.12	6.80	7.91	7.47	7.75	7.11	8.19

Barre. Most of the children were white. The range of nationalities was extensive; Italians, Poles, Lithuanians, Slovaks, Germans and Russian Jews were predominant. Scores were tabulated according to age. (See Table I.) The means for the successive ages were accepted as age norms.

Results.—Correlations with following intelligence tests were:

Pintner-Cunningham Primary Mental Test (kindergarten and first grade)	.75
Otis Primary Test (second to fifth grades)	.72
Henmon-Nelson Test (sixth and seventh grades)	.69

The present study seems to indicate that copying of drawings does not test to any great extent what is usually known as "drawing ability." To copy geometrical figures requires an appreciation of spatial relationships and eye-hand coördination. Both these are closely connected—or perhaps are identical with the so-called intelligence tested in the regular tests of mental ability. Terman[1] and Goodenough[2] came to the same conclusions in their respective studies.

SUMMARY AND CONCLUSIONS

(1) The purpose of the study was: (1) to standardize a series of drawings of increasing difficulty and (2) to ascertain the degree of correlation between the ability to copy the drawings and general intelligence.

(2) A scale was constructed. This scale consists of twenty drawings arranged in order of difficulty. The sequence of these drawings was established on the basis of the copyings made by three hundred children between the ages of four and twelve.

(3) The results indicate that scales of similar type may be used as a test of intellectual development. While they cannot be substituted for the more elaborate intelligence scales and tests, they can be of great value where tentative classifications are imperative. The tests were readministered to approximately three hundred children in the lower grade and a retest coefficient of .87 $\pm$ .03 was found between the two administrations. It would seem, therefore, that the test is reliable for diagnosing both mentally retarded and bright children.

(4) The average correlation with Otis Primary Test, Pintner-Cunningham Primary Mental Test and Henmon-Nelson Test of Mental Ability (third to eighth grade) is .72 with average PE of .02.

---

[1] *Op. cit.*, p. 157. The figure must be perceived as a whole, not simply as a group of meaningless lines. In the second place, success depends upon the ability to use visual impression in guiding a rather complex set of motor coördinations.

[2] Goodenough, Florence, L.: *Measurement of Intelligence by Drawings*, 1926, p. 82.

(5) The average correlation with teachers' judgment of mental ability was .85 in the kindergarten and the first grade, .75 in the second grade, .73 in the third grade. The correlations in the higher grades were less significant.

(6) The drawings are easily administered and scored. This is a great convenience.

BOOK REVIEWS

J. F. DASHIELL. *Fundamentals of General Psychology.* Boston: Houghton Mifflin Company, 1937, pp. XXIII + 655.

This book is in a sense a revision of the author's earlier text of 1928. There is so much change in both material and point of view, however, that the present treatise may be considered as a new book. The author has moved in presenting his materials from a rather strict behaviorism in 1928 to a much broader viewpoint in which account is taken of the contributions of not only objectivists, but of the Gestaltists and others. The stand is taken that psychology can be presented as a scientific body of knowledge and at the same time help the student in his orientation and practical problems.

There are many features in the book which make it a particularly attractive text: (1) The presentation is thorough, well balanced, and critical. Satisfactory emphasis and space have been given to practically all parts of the discussion. (2) There is a generous citation of experimental data. While some readers may consider that these citations are too frequent, the author has evaluated the data and made them an integral part of his discussion in a remarkably successful manner. (3) The figures are numerous, excellent, and to the point. It is especially gratifying to find a *complete* explanation accompanying each figure although this descriptive material frequently runs to one-fourth of a page. (4) The discussions are enriched by frequent citation of examples from everyday life. (5) The developmental approach is introduced at appropriate places in the discussion.

Although all chapters are of high quality, some sections should receive special mention: The section on motivation is, perhaps, the best available in any elementary text. Other exceptional chapters include those on emotion, learning, social behavior, and thinking.

The critical reader will, of course, have reservations concerning certain materials and methods in this book. Those which occur to the reviewer follow: (1) Some of the material is presented too briefly or is too technical to be read with profit by the elementary student. Examples are the mathematical formulation of a rat's orientation to an incline plane, discussion of chronaxie, tetrad differences, factor analysis, and typological psychology. (2) It seems reasonable to suggest that whatever statistical method is introduced into an elementary text should be the simplest that will serve to illustrate the principle at hand. The group method of computing measures of

central tendency and use of the product-moment correlation, therefore, seem out of place. (3) Some readers will feel that undue space and emphasis is devoted to the physiological aspects of motor functions and to neural organization. It must be admitted, however, that these sections fit nicely into the pattern of the book. (4) Recent studies reveal that eight per cent of men are partially color blind rather than four as cited. (5) Small print is employed too frequently and too extensively for textual material other than footnotes.

It is easy to predict that this excellent book will be received by instructors with a great deal of enthusiasm and that it will be widely used. It will undoubtedly turn out to be the "book of the year" among texts in elementary psychology. In fact, many will consider it the best introductory text that has appeared for several years.

University of Minnesota. MILES A. TINKER.

NORMA V. SCHEIDEMANN. *The Psychology of Exceptional Children.* Volume II. Boston: Houghton Mifflin Co., 1937, pp. XII + 460.

After an interval of six years Dr. Scheidemann publishes the book being reviewed as a second volume of her *Psychology of Exceptional Children.* In the two volumes a wide variety of exceptional children are discussed, but the chapters of both volumes, and more particularly the present one, give no evidence of organization. The second volume is really a collection of essays on ten types of children who can be considered exceptional. It is implied in the preface that this volume is to consider the physical types of exception and, therefore, the initial chapter is devoted to an elementary description of the nervous system. Why the nervous system was singled out for special discussion when only three of the types discussed are directly concerned with neurological deviations is an unanswered question.

The separate chapters are devoted to the following subjects: cerebral birth injuries, post-encephalitis, enuresis, albinism, preschizophrenia, puberty praecox, premature birth, bilingualism, eidetic imagery, and congenital syphilis. This list is in the same order as the chapters and quite evidently it represents no plan of organization. Each of the essays is, however, a fairly adequate discussion of its problem. From the preface one gathers that the book is written for teachers, and in the separate chapters there is some discussion of the educational implications of the deviation. While there is a glossary of technical terms one might question if the material is presented in

such a manner that it will be understood by teachers without a great deal of additional background reading. C. M. LOUTTIT.
Indiana University.

MARTHA C. HARDY and CAROLYN H. HOEFER. *Healthy Growth.* Chicago: University of Chicago Press, 1936, pp. XII + 360.

The findings of an investigation relative to the value of a health instruction program in the public schools, together with a discussion of the relationships between physical well-being and other aspects of child development, form the subject-matter of this book. The investigators have attempted to secure a comprehensive view of the whole child by means of a longitudinal study of a group of children. Although the authors claim that the same children were followed over a twelve-year period, the data which they report cover only the four-year period from 1923 to 1927. Records of physical, mental, and educational growth, and behavior ratings by parents and teachers were obtained for each child over this period, and, in addition, interest, attitude, and personality questionnaires were filled out annually by the children.

This extensive investigation by Hardy and the late Carolyn Hoefer was made possible by the coöperation of the school authorities of Joliet, Ill., and the financial assistance of the Elizabeth McCormick Memorial Fund. The participants were four hundred nine pupils in the Joliet schools, divided into three groups. The so-called "experimental" group took part in a program of health education, consisting of regular classroom instruction, yearly medical examinations, consultations with child specialists, and follow-up work by nurses. The control group's program was similar, except that they received no classroom instruction. A supplementary group was composed of children, who, because of transfer from one school to another, received classroom instruction for less than the entire experimental period. The experimental and control groups were supposedly equated at the start of the experiment with regard to age, intelligence, initial health status, and several other traits. The equating, however, was far from satisfactory, particularly in the case of socio-economic status and intelligence, and only by courtesy can the groups be said to be matched. For certain sub-analyses of the results, the authors have attempted to eliminate some of the errors which might arise from this lack of equality by selecting more carefully matched groups; this procedure, however, reduces the number of cases on which their findings are based, and does not fully compensate for the initial lack of equality between the groups.

It was found that a decided improvement in the health of both groups, as measured by the physicians' yearly ratings, took place during the period of instruction. The group which had the most extensive contact with the health program, *i.e.*, the experimental or class-instructed group, made the greatest gain, but the differences between the groups were probably not significant. With regard to rate of physical growth and development, the findings were likewise suggestive of a more rapid advance for the experimental group.

With reference to rate of mental growth, a study of the yearly performance of the children on the Stanford-Binet and on a modified form of the Pintner-Paterson Performance Scale revealed an insignificant superiority for the experimental group. Certain departures from the standard techniques of administration for both of these tests render the findings even less conclusive. School progress and educational achievement likewise failed to show any association with participation in the health-education program. No differences were found between the groups with regard to extra-curricular interests, attitudes, or behavior adjustment.

The two final chapters deal with the importance of health in general during childhood. An analysis of the records of children classified as in "very good," "fair" or "very poor" health indicates strikingly that the healthier children tend to be brighter, better adjusted to their environment, both in school and out; that they have an accelerated rate of physical growth, and make more rapid school progress than their less healthy classmates. Significant positive correlations between health status and measures of educational achievement, intelligence, personality, and a number of other traits indicate the all-round superiority of the children in good health and, conversely, the inferiority of the children in poor health. The authors do not state explicitly that they consider health to be a causative factor in this superiority, but such a belief is strongly implied throughout the chapter.

Unfortunately, the findings of the investigation are vitiated to a great extent by several factors. The poor equating of the groups has already been mentioned. Most of the measures employed in the study—physicians' ratings, estimates of socio-economic status, interest and attitude questionnaires, personality ratings, and the Performance Scale—are of unknown or disappointing reliability. The variable which was being studied, *viz.*, classroom instruction in health topics, varied from class to class, depending on the teacher and the individual needs of each class.

The tables in which the data have been summarized are not always clear, and in some instances, *e.g.*, Tables 20 and 50, are marked by computational inaccuracy and erroneous headings. Critical ratios, which have been computed for almost all the group comparisons in the study, are sometimes given only for the difference between the initial and final status of each group, rather than for the differences in gains made by the two groups. The topical and author index is quite inadequate. It is regrettable that such a broadly conceived testing program, with such possibilities for a thorough developmental study of the whole child, did not yield more satisfactory results. In fairness to the authors, however, it should be pointed out that the investigation was planned more than fifteen years ago, and they were forced to employ less adequate measures, tests, and techniques than are available at present.

Despite the above mentioned shortcomings, the authors present a wealth of information in this volume which is deserving of careful consideration by those interested in child health and development.

ROGER T. LENNON.

Fordham University.

MELVIN E. HAGGERTY. *The Evaluation of Higher Institutions. Volume II, The Faculty.* Chicago: University of Chicago Press, 1937, pp. IX + 218.

This carefully-written report of the Committee on Revision of Standards appointed by the Commission on Institutions of Higher Education of the North Central Association is deserving of careful study by all college and university teachers. Faced with the necessity of developing an improved method for accrediting institutions of higher education, the committee analyzed a great volume of data from fifty-seven schools seeking factors which might be related to general institutional excellence. The criterion of excellence most generally used was a composite involving (*a*) ratings by four experts, (*b*) student achievement examination scores in reading, social studies, natural science, French and German, (*c*) estimated effectiveness of instruction in mathematics, and (*d*) rankings of each institution in terms of the success its graduates achieved in graduate schools.

The institutional characteristics named below are followed by a rank coefficient of correlation indicating their relationship to this criterion:

1. Faculty competence	+.45	6. Library	+.64
2. Faculty organization	+.67	7. Induction of students	+.60
3. Conditions of service	+.56	8. Student personnel	+.57
4. Curriculum	+.63	9. Administration	+.60
5. Instruction	+.65	10. Plant	+.61
		11. Finance	+.47

A weighted mean square contingency coefficient between a composite of the eleven characteristics given above and the criterion was $+.79 \pm .03$.

As a result of their inquiry the committee recommends the abolition of fixed standards as heretofore employed and the adoption instead of a pattern map. This pattern map "has proved helpful to the investigators, to the association in its accrediting activities, and to the institutions to whom reports have been made." STEPHEN M. COREY.

University of Wisconsin.